# Characterizing Classes of Regular Languages Using Prefix Codes of Bounded Synchronization Delay

**Volker Diekert[1] and Tobias Walter[*†2]**

1 **University of Stuttgart, FMI, Stuttgart, Germany**
`diekert@fmi.uni-stuttgart.de`

2 **University of Stuttgart, FMI, Stuttgart, Germany**
`walter@fmi.uni-stuttgart.de`

## Abstract

In this paper we continue a classical work of Schützenberger on codes with bounded synchronization delay. He was interested in characterizing those regular languages where the groups in the syntactic monoid belong to a variety $\mathbf{H}$. He allowed operations on the language side which are union, intersection, concatenation and modified Kleene-star involving a mapping of a prefix code of bounded synchronization delay to a group $G \in \mathbf{H}$, but no complementation. In our notation this leads to the language classes $\mathrm{SD}_G(A^\infty)$ and $\mathrm{SD}_{\mathbf{H}}(A^\infty)$. Our main result shows that $\mathrm{SD}_{\mathbf{H}}(A^\infty)$ always corresponds to the languages having syntactic monoids where all subgroups are in $\mathbf{H}$. Schützenberger showed this for a variety $\mathbf{H}$ if $\mathbf{H}$ contains Abelian groups, only. Our method shows the general result for all $\mathbf{H}$ directly on finite and infinite words. Furthermore, we introduce the notion of *local Rees extensions* which refers to a simple type of classical Rees extensions. We give a decomposition of a monoid in terms of its groups and local Rees extensions. This gives a somewhat similar, but simpler decomposition than in Rhodes' synthesis theorem. Moreover, we need a singly exponential number of operations, only. Finally, our decomposition yields an answer to a question in a recent paper of Almeida and Klíma about varieties that are closed under Rees extensions.

**1998 ACM Subject Classification** F.4.3 Formal Languages

**Keywords and phrases** formal language, synchronization delay, variety, Rees extension

**Digital Object Identifier** 10.4230/LIPIcs.ICALP.2016.129

*In memoriam: Marcel-Paul Schützenberger (1920–1996)*

## 1 Introduction

A fundamental result of Schützenberger characterizes the class of star-free languages SF as exactly those languages which are group-free, that is, aperiodic [15]. One usually abbreviates this result by $\mathrm{SF} = \mathbf{AP}$. Schützenberger also found another, but less prominent characterization of SF: the star-free languages are exactly the class of languages which can be defined inductively by finite languages and closure under finite union, concatenation, and the Kleene-star restricted to prefix codes of bounded synchronization delay [17]. This result

* Supported by the German Research Foundation (DFG) under grant DI 435/6-1.
† The authors thank an anonymous referee for the suggestion of the notation *group-controlled star*.


43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016).
Editors: Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi; Article No. 129; pp. 129:1–129:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

is abbreviated by $\mathbf{AP} = \mathrm{SD}$. It is actually stronger than the famous $\mathrm{SF} = \mathbf{AP}$ because $\mathrm{SD} \subseteq \mathrm{SF} \subseteq \mathbf{AP}$ is relatively easy, see [11, Chapter VIII], so $\mathrm{SF} = \mathbf{AP}$ follows from $\mathbf{AP} \subseteq \mathrm{SD}$. The extension $\mathrm{SF} = \mathbf{AP}$ to infinite words is due to Perrin [10]. The result $\mathbf{AP} = \mathrm{SD}$ for infinite words was obtained much later in [5]. It became possible thanks to a "local divisor approach", which also is a main tool in this paper.

Schützenberger did not stop by showing $\mathbf{AP} = \mathrm{SD}$. In retrospective he started a program: in [16] he was able to prove an analogue of $\mathbf{AP} = \mathrm{SD}$ for languages where syntactic monoids have Abelian subgroups, only. In our notation $\mathbf{AP} = \mathrm{SD}$ means $\overline{\mathbf{1}}(A^\infty) = \mathrm{SD}_{\mathbf{1}}(A^\infty)$; and the main result in [16] is "essentially" equivalent to $\overline{\mathbf{Ab}}(A^*) = \mathrm{SD}_{\mathbf{Ab}}(A^*)$. (We write "essentially" because using the structure theory of Abelian groups, a sharper version than $\overline{\mathbf{Ab}}(A^*) = \mathrm{SD}_{\mathbf{Ab}}(A^*)$ is possible.) The proofs [16] use deep results in semigroup theory; and no such result beyond Abelian groups was known so far. Our result generalizes $\overline{\mathbf{Ab}}(A^\infty) = \mathrm{SD}_{\mathbf{Ab}}(A^\infty)$ to every variety $\mathbf{H}$ of finite groups: we show $\overline{\mathbf{H}}(A^\infty) = \mathrm{SD}_{\mathbf{H}}(A^\infty)$. We were able to prove it with much less technical machinery compared to [16]. For example, no knowledge in Krohn-Rhodes theory is required.

Actually, our result is a generalization of $\overline{\mathbf{Ab}}(A^*) = \mathrm{SD}_{\mathbf{Ab}}(A^*)$ [16] and also of $\mathbf{AP}(A^\infty) = \mathrm{SD}(A^\infty)$ [5]. More precisely, we give a characterization of languages which are recognized by monoids where all subgroups belong to $\mathbf{H}$. The characterization uses an inductive scheme starting with all finite subsets of finite words, allows concatenation, finite union, no (!) complementation, but a restricted use of a group-controlled star (resp. group-controlled $\omega$-power). Let us explain the *group-controlled star* in our context. Instead of putting the star above a single language, consider first a disjoint union $K = \bigcup \{K_g \mid g \in G\}$ where $G$ is a finite group and each $K_g$ is regular in $A^*$. The "group-controlled star", more precisely the "$G$-controlled star", associates with such a disjoint union the following language:

$$\{u_{g_1} \cdots u_{g_k} \in K^* \mid u_{g_i} \in K_{g_i} \wedge g_1 \cdots g_k = 1 \in G\}.$$

Clearly, we obtain a regular language, but without any restriction, allowing such a "group star" yields all regular languages, even in the case of the trivial group. So, the construction is of no interest without a simultaneous restriction. The restriction considered in [16] yields an inductive scheme to define a class $\mathcal{C}$. The restriction says that such a group-controlled star is allowed only over a disjoint union $K = \bigcup \{K_g \mid g \in G\}$ where each $K_g$ already belongs to $\mathcal{C}$ and where $K$ is, in addition, a prefix code of bounded synchronization delay. The initials in "synchronization delay" led to the notation SD; and an indexed version $\mathrm{SD}_G$ (resp. $\mathrm{SD}_{\mathbf{H}}$) refers to *synchronization delay* over $G$ (resp. over a finite group in $\mathbf{H}$). Since we also deal with infinite words we apply the same restriction to $\omega$-powers.

Our results give also a new characterization for various other classes. For example, by a result of Straubing, Thérien and Thomas [20], the class of languages, having syntactic monoids where all subgroups are solvable, coincides with $(\mathrm{FO} + \mathrm{MOD})[<]$. Here, $(\mathrm{FO} + \mathrm{MOD})[<]$ means the class of languages defined by the logic $(\mathrm{FO} + \mathrm{MOD})[<]$. Thus, we are able to give a new language characterization: $(\mathrm{FO} + \mathrm{MOD})[<](A^\infty) = \mathrm{SD}_{\mathbf{Sol}}(A^\infty)$.

Moreover, as a sort of byproduct of $\overline{\mathbf{H}} = \mathrm{SD}_{\mathbf{H}}$, we obtain a simple and purely algebraic characterization of the monoids in $\overline{\mathbf{H}}$. Every monoid in $\overline{\mathbf{H}}$ can be decomposed in at most exponentially many iterated Rees extensions of groups in $\mathbf{H}$. The iteration uses only a very restricted version of Rees extensions: *local Rees extensions*. This means we obtain every finite monoid which is not a group as a divisor of a Rees extension between two proper divisors of $M$, one of them a proper submonoid, the other one a "local divisor".

Our decomposition result is similar to the synthesis theory of Rhodes and Allen [13]. Moreover, it yields a singly exponential bound on the number of operations whereas no such bound was known by [13]. Finally, using this decomposition, we answer a recent question of Almeida and Klíma [1] concerning varieties which are closed under Rees extensions.

## 2 Preliminaries

Throughout, $A$ denotes a finite alphabet and $A^*$ is the free monoid over $A$. It consists of all finite words. The empty word is denoted by 1 as the neutral elements in other monoids or groups. The set of non-empty finite words is $A^+$; it is the free semigroup over $A$. By $A^\omega$ we denote the set of all infinite words with letters in $A$. For a set $K \subseteq A^*$, we let $K^\omega = \{u_1 u_2 \cdots \mid u_i \in K, u_i \text{ non-empty}, i \in \mathbb{N}\} \subseteq A^\omega$. In particular, $K^\omega = (K \setminus \{1\})^\omega$. Since our results concern finite and infinite words, it is convenient to treat finite and infinite words simultaneously. We define $A^\infty = A^* \cup A^\omega$ to be the set of finite or infinite words. Accordingly, a *language* $L$ is a subset of $A^\infty$. We say that $L$ is *regular*, if first, $L \cap A^*$ is regular and second, $L \cap A^\omega$ is $\omega$-regular in the standard meaning of formal language theory. In order to study regular languages algebraically, one considers finite monoids. A *divisor* of a monoid $M$ is a monoid $N$ which is a homomorphic image of a subsemigroup of $M$. In this case we write $N \preceq M$. A subsemigroup $S$ of $M$ is in our setting a divisor if and only if $S$ is a monoid (but not necessarily a submonoid of $M$). A *variety* of finite monoids – hence, in Birkhoff's setting: a *pseudovariety* – is a class of finite monoids $\mathbf{V}$ which is closed under finite direct products and under division:

- If $I$ is a finite index set and $M_i \in \mathbf{V}$ for each $i \in I$, then $\prod_{i\in I} M_i \in \mathbf{V}$. In particular, the trivial group $\{1\}$ belongs to $\mathbf{V}$.
- If $M \in \mathbf{V}$ and $N \preceq M$, then $N \in \mathbf{V}$.

Classical formal language theory states "regular" is the same as "recognizable". This means: $L \subseteq A^*$ is regular if and only if its syntactic monoid is finite; $L \subseteq A^\omega$ is regular if and only if its syntactic monoid in the sense of Arnold [2] is finite and, in addition, $L$ is saturated by the syntactic congruence, see eg. [11, 21]. Here we use a notion of recognizability which applies to languages $L \subseteq A^\infty$. Let $\varphi : A^* \to M$ be a homomorphism to a finite monoid $M$. First, we define a relation $\sim_\varphi$ as follows. If $u \in A^*$ is a finite word, then we write $u \sim_\varphi v$ if $v$ is finite and $\varphi(u) = \varphi(v)$. If $u \in A^\omega$ is an infinite word, then we write $u \sim_\varphi v$ if $v$ is infinite and if there are factorizations $u = u_1 u_2 \cdots$ and $v = v_1 v_2 \cdots$ into finite nonempty words such that $\varphi(u_i) = \varphi(v_i)$ for all $i \geq 1$. It is easy to see that $\sim_\varphi$ is not transitive on infinite words, in general. Therefore, we consider its transitive closure $\approx_\varphi$. If $u, v \in A^*$, then we have

$$u \sim_\varphi v \iff u \approx_\varphi v \iff \varphi(u) = \varphi(v).$$

If $\alpha, \beta \in A^\omega$, then we have $\alpha \approx_\varphi \beta$ if and only if there is sequence of infinite words $\alpha_0, \ldots \alpha_k$ such that

$$\alpha = \alpha_0 \sim_\varphi \cdots \sim_\varphi \alpha_k = \beta.$$

We say that $L \subseteq A^\infty$ is *recognizable* by $M$ if there exists a homomorphism $\varphi : A^* \to M$ such that $u \in L$ and $u \sim_\varphi v$ implies $v \in L$. We also say that $M$ or $\varphi$ recognizes $L$ in this case. The connection to the classical notation is as follows. A regular language $L \subseteq A^\infty$ is recognizable (in our sense) by $\varphi$ if and only if the syntactic monoids of $L \cap A^*$ and $L \cap A^\omega$ are divisors of $M$. Another equivalent definition can be given in terms of Wilke algebras [22].

Every variety $\mathbf{V}$ defines a family of regular languages $\mathbf{V}(A^\infty)$ as follows: we let $L \in \mathbf{V}(A^\infty)$ if there exists a monoid $M \in \mathbf{V}$ which recognizes $L$. Further, we define $\mathbf{V}(A^*) = \{L \subseteq A^* \mid L \in \mathbf{V}(A^\infty)\}$ and $\mathbf{V}(A^\omega) = \{L \subseteq A^\omega \mid L \in \mathbf{V}(A^\infty)\}$ . A variety of finite groups is a variety of finite monoids which contains only groups. Throughout $\mathbf{H}$ denotes a variety of finite groups. Special cases are the varieties

- $\mathbf{1}$: the trivial group $\{1\}$, only.
- $\mathbf{Ab}$: all finite Abelian groups.
- $\mathbf{Sol}$: all finite solvable groups.
- $\mathbf{Sol}_q$: all finite solvable groups where the order is divisible by some power of $q$.
- $\mathbf{G}$: all finite groups.

According to standard notation $\overline{\mathbf{H}}$ denotes the variety of finite monoids where all subgroups belong to $\mathbf{H}$. It is not completely obvious, but a classical fact [9], that $\overline{\mathbf{H}}$ is indeed a variety. In fact, it is the maximal variety $\mathbf{V}$ such that $\mathbf{V} \cap \mathbf{G} = \mathbf{H}$.

Clearly, $\overline{\mathbf{G}}$ is the class of all finite monoids. The most prominent subclass is $\overline{\mathbf{1}}$: it is the variety of aperiodic monoids $\mathbf{AP}$. The class $\mathbf{AP}(A^\infty) = \overline{\mathbf{1}}(A^\infty)$ admits various other characterizations as subsets of $A^\infty$. For example, it is the class of star-free languages $\mathrm{SF}(A^\infty)$, it is the class of first-order definable languages, and it is the class of definable languages in linear temporal logic over finite or infinite words: $\mathrm{LTL}(A^\infty)$ .

**Local divisors.** Let $M$ be a finite monoid and $c \in M$. Consider the set $cM \cap Mc$ with a new multiplication $\circ$ which is defined as follows:

$$mc \circ cn = mcn.$$

A straightforward calculation shows that $cM \cap Mc$ becomes a monoid with this operation where the neutral element of $M_c$ is $c$. Thus, the structure $M_c = (cM \cap Mc, \circ, c)$ defines a monoid. We say that $M_c$ is the *local divisor* of $M$ at $c$. If $c$ is a unit, then $M_c$ is isomorphic to $M$. If $c = c^2$, then $M_c$ is the standard "local monoid" at the idempotent $c$.

The important fact is that $M_c$ is always a divisor of $M$ and that $|M_c| < |M|$ as soon as $c$ is not a unit of $M$. Indeed, the mapping $\lambda_c : \{x \in M \mid cx \in Mc\} \to M_c$ given by $\lambda_c(x) = cx$ is a surjective homomorphism. Moreover, if $c$ is not a unit, then $1 \notin cM \cap Mc$, hence $|M_c| < |M|$. Thus, if $M$ belongs to some variety $\mathbf{V}$, then $M_c$ belongs to the same variety. If $M$ is not a group, then we find some nonunit $c \in M$ and the local divisor $M_c$ is smaller than $M$. This makes the construction useful for induction. For a survey on the local divisor technique we refer to [6].

**Rees extensions.** Let $N, L$ be monoids and $\rho : N \to L$ be any mapping. The *Rees extension* $\mathrm{Rees(N, L, \rho)}$ is a classical construction for monoids [12, 14], frequently described in terms of matrices. Here, we use an equivalent definition as in [7]. As a set we define

$$\mathrm{Rees(N, L, \rho)} = N \cup (N \times L \times N)\,.$$

The multiplication $\cdot$ on $\mathrm{Rees(N, L, \rho)}$ is given by

$$\begin{aligned} n \cdot n' &= nn' && \text{for } n, n' \in N, \\ n \cdot (n_1, m, n_2) \cdot n' &= (nn_1, m, n_2 n') && \text{for } n, n', n_1, n_2 \in N, m \in L, \\ (n_1, m, n_2) \cdot (n_1', m', n_2') &= (n_1, m\rho(n_2 n_1') m', n_2') && \text{for } n_1, n_1', n_2, n_2' \in N, m, m' \in L. \end{aligned}$$

The neutral element of $\mathrm{Rees(N, L, \rho)}$ is $1 \in N$ and $N \subseteq \mathrm{Rees(N, L, \rho)}$ is an embedding of monoids. In general, $L$ is not a divisor of $\mathrm{Rees(N, L, \rho)}$. The following property holds.

▶ **Lemma 1.** *Let $N \preceq N'$ and $L \preceq L'$. Given $\rho : N \to L$, there exists a mapping $\rho' : N' \to L'$ such that* $\mathrm{Rees(N, L, \rho)}$ *is a divisor of* $\mathrm{Rees(N', L', \rho')}$.

**Proof.** First, assume that $N$ (resp. $L$) is submonoid in $N'$ (resp. $L'$). Let $\rho' : N' \to L'$ be any function such that $\rho'|_N = \rho$. The mapping $\pi : \mathrm{Rees(N, L, \rho)} \to \mathrm{Rees(N', L', \rho')}$ given by $\pi(n) = n$ and $\pi(n_1, \ell, n_2) = (n_1, \ell, n_2)$ is an injective homomorphism.

Second, let $\varphi : N' \to N$ and $\psi : L' \to L$ be surjective homomorphisms. Let $\rho' : N' \to L'$ be a function such that $\rho'(n) \in \psi^{-1}(\rho(\varphi(n)))$. Let $\pi : \mathrm{Rees(N', L', \rho')} \to \mathrm{Rees(N, L, \rho)}$ be the mapping defined by $\pi(n) = \varphi(n)$ and $\pi(n_1, \ell, n_2) = (\varphi(n_1), \psi(\ell), \varphi(n_2))$. It is clear that $\pi$ is surjective. It is a homomorphism since

$$\pi((n_1, \ell, n_2) \cdot (n_1', \ell', n_2')) = \pi(n_1, \ell\rho'(n_2 n_1')\ell', n_2') = (\varphi(n_1), \psi(\ell)\underbrace{\psi(\rho'(n_2 n_1'))}_{=\rho(\varphi(n_2 n_1'))}\psi(\ell'), \varphi(n_2'))$$
$$= (\varphi(n_1), \psi(\ell), \varphi(n_2)) \cdot (\varphi(n_1'), \psi(\ell'), \varphi(n_2')) = \pi(n_1, \ell, n_2) \cdot \pi(n_1', \ell', n_2'). \qquad \blacktriangleleft$$

We are mainly interested in the case where $N$ and $L$ are proper divisors of a given finite monoid $M$. This leads to the notion of local Rees monoids. More precisely, let $M$ be a finite monoid, $N$ be a proper submonoid of $M$ and $M_c$ be a local divisor of $M$ at $c$ where $c$ is not a unit. The *local Rees extension* $\mathrm{LocRees(N, M_c)}$ is defined as the Rees extension $\mathrm{Rees(N, M_c, \rho_c)}$ where $\rho_c$ denotes the mapping $\rho_c : N \to M_c; x \mapsto cxc$.

For a variety $\mathbf{V}$ we define $\mathrm{Rees}(\mathbf{V})$ to be the least variety which contains $\mathbf{V}$ and is closed under taking Rees extensions and $\mathrm{LocRees}(\mathbf{V})$ to be the least variety which contains $\mathbf{V}$ and is closed under local Rees extensions.

## 2.1 Schützenberger's SD classes

Schützenberger gave a language theoretical characterization of the class of star-free languages $\mathrm{SF}(A^*)$ avoiding complementation, but allowing the star-operation to prefix codes of bounded synchronization delay [17].

A language $K \subseteq A^+$ is called *prefix code* if it is *prefix-free*. That is: $u, uv \in K$ implies $u = uv$. A prefix-free language $K$ is a code since every word $u \in K^*$ admits a unique factorization $u = u_1 \cdots u_k$ with $k \geq 0$ and $u_i \in K$. Note that the empty set $\emptyset$ is considered to be a prefix code. More generally, if $L \subseteq A^+$ is any subset, then $K = L \setminus LA^+$ is a prefix code. A prefix code $K$ has *bounded synchronization delay* if for some $d \in \mathbb{N}$ and for all $u, v, w \in A^*$ we have: if $uvw \in K^*$ and $v \in K^d$, then $uv \in K^*$. Note that the condition implies that for all $uvw \in K^*$ with $v \in K^d$, we have $w \in K^*$, too. If $d$ is given explicitly, $K$ is said to have synchronization delay $d$. Every subset $B \subseteq A$ (including the empty set) yields a prefix code with synchronization delay 0. If we have $c \in A \setminus B$, then $B^*c$ is a prefix code with synchronization delay 1. If $K$ is any prefix code with (or without) bounded synchronization delay, then $K^m$ is a prefix code for all $m \in \mathbb{N}$, but for $m \geq 2$ it is never of bounded synchronization delay.

Consider a disjoint union $K = \bigcup\{K_g \mid g \in G\}$ of a prefix code $K$ with bounded synchronization delay where $G$ is a finite group and each $K_g$ is regular in $A^*$. The *$G$-controlled star* associates with such a disjoint union the following language:

$$\{u_{g_1} \cdots u_{g_k} \in K^* \mid u_{g_i} \in K_{g_i} \wedge g_1 \cdots g_k = 1 \in G\}.$$

Another view of the $G$-controlled star of $K$ is the following: Let $\gamma_K : K \to G$ be a mapping such that $K_g = \gamma_K^{-1}(g)$ and let $\gamma : K^* \to G$ denote the canonical extension of $\gamma_K$ to a homomorphism from the free submonoid $K^* \subseteq A^*$ to $G$, then the $G$-controlled star of $K$ is exactly the set $\gamma^{-1}(1)$. The generalization to infinite words $\gamma^{-1}(1)^\omega$ is called the $G$-controlled $\omega$-power. Let $\mathcal{C}$ be a class of languages. We say that $\mathcal{C}$ is closed under $G$-controlled star ($\omega$-power) if $K$ is a prefix code with bounded synchronization delay, $K_g \in \mathcal{C}$ for all $g \in G$, then the $G$-controlled star $\gamma^{-1}(1)$ ($\omega$-power $\gamma^{-1}(1)^\omega$) is in $\mathcal{C}$. For a variety of groups $\mathbf{H}$ we say that $\mathcal{C}$ is closed under $\mathbf{H}$-controlled star ($\omega$-power) if $\mathcal{C}$ is closed under $G$-controlled star ($\omega$-power) for every group $G \in \mathbf{H}$. By $\mathrm{SD}_G(A^\infty)$ we denote the smallest class of regular languages such that $\emptyset \in \mathrm{SD}_G(A^\infty)$, $\{a\} \in \mathrm{SD}_G(A^\infty)$ for all letters $a \in A$, $\mathrm{SD}_G(A^\infty)$ is closed under finite union and concatenation, i.e., $L, K \in \mathrm{SD}_G(A^\infty)$ implies $L \cup K$ and $(L \cap A^*) \cdot K$ are both in $\mathrm{SD}_G(A^\infty)$, and $\mathrm{SD}_G(A^\infty)$ is closed under $G$-controlled star and $G$-controlled $\omega$-power. We also define

$$\mathrm{SD}_G(A^*) = \{L \subseteq A^* \mid L \in \mathrm{SD}_G(A^\infty)\} \quad \text{and} \quad \mathrm{SD}_G(A^\omega) = \{L \subseteq A^\omega \mid L \in \mathrm{SD}_G(A^\infty)\}\,.$$

Note that for every homomorphism $\gamma : A^* \to G$ we have $\gamma^{-1}(1) \in \mathrm{SD}_G(A^*)$ and $\gamma^{-1}(1)^\omega \in \mathrm{SD}_G(A^\omega)$. This follows because first, $A$ is a prefix code of bounded synchronization delay and second, all finite subsets of $A$ are in $\mathrm{SD}_G(A^*)$.

Unlike the case of star-free sets, the definition of $\mathrm{SD}_G(A^\infty)$ does not use any complementation. By induction: for $L \subseteq A^\infty$ we have $L \in \mathrm{SD}_G(A^\infty)$ if and only if we can write $L = L_1 \cup L_2$ with $L_1 \in \mathrm{SD}_G(A^*)$ and $L_2 \in \mathrm{SD}_G(A^\omega)$. In the special case where $G = \{1\}$ is the trivial group, we also simply write SD instead of $\mathrm{SD}_{\{1\}}$. In this case closure under $\{1\}$-controlled stars ($\omega$-powers) can be rephrased in simpler terms as follows: If $K \in \mathrm{SD}(A^*)$ is a prefix code of bounded synchronization delay, then $K^* \in \mathrm{SD}(A^*)$ and $K^\omega \in \mathrm{SD}(A^\omega)$.

In [16] Schützenberger showed (using a different notation) $\mathrm{SD}_\mathbf{H}(A^*) \subseteq \overline{\mathbf{H}}(A^*)$, but the converse only for $\mathbf{H} \subseteq \mathbf{Ab}$, see Proposition 6 for the first inclusion. Our aim is to show $\overline{\mathbf{H}}(A^\infty) \subseteq \mathrm{SD}_\mathbf{H}(A^\infty)$ for all $\mathbf{H}$, cf. Theorem 4. We begin with a technical lemma.

▶ **Lemma 2.** *Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay and let $\gamma : K^* \to G$ be a homomorphism such that $\gamma^{-1}(g) \cap K \in \mathrm{SD}_G(A^*)$ for all $g \in G$, then we have $\gamma^{-1}(g) \in \mathrm{SD}_G(A^*)$ for all $g \in G$.*

**Proof.** For a word $w = u_1 \cdots u_k \in K^*$ we define $P(w) = \{\gamma(u_1 \cdots u_i) \mid 1 \le i \le k\} \subseteq G$ to be the set of prefixes of $w$ in $G$. By an induction on $|P(w)|$ we construct languages $L(w) \in \mathrm{SD}_G(A^*)$ such that $w \in L(w) \subseteq \gamma^{-1}(\gamma(w))$ and the number $|\{L(w) \mid w \in K^*\}|$ of such languages is finite. The base case $|P(w)| = 0$ implies $g = 1$. We may choose $L(w) = \gamma^{-1}(1)$ and we are done, since $\gamma^{-1}(1) \in \mathrm{SD}_G(A^*)$ by definition. Hence, we may assume $|P(w)| \ge 1$. Let $g_1 = \gamma(u_1)$ and choose $i$ maximal such that $g_1 = \gamma(u_1 \cdots u_i)$. Then we have $u_1 \cdots u_i \in (K \cap \gamma^{-1}(g_1)) \cdot \gamma^{-1}(1)$. Note that $P(w') = g_1^{-1} \cdot \{\gamma(u_1 \cdots u_j) \mid i < j \le k\}$ for $w' = u_{i+1} \cdots u_k$. By choice of $i$ we have $g_1 \notin \{\gamma(u_1 \cdots u_j) \mid i < j \le k\}$ and therefore $|P(w')| = |\{\gamma(u_1 \cdots u_j) \mid i < j \le k\}| < |P(w)|$. By induction there exists $L(w')$ and we let $L(w) = (K \cap \gamma^{-1}(g_1)) \cdot \gamma^{-1}(1) \cdot L(w')$. The number of $|\{L(w) \mid w \in K^*\}|$ is therefore bounded by $\sum_{i=0}^{|G|} |G|^i$ which is less than $|G|^{|G|+1}$. The result follows because we can write $\gamma^{-1}(g) = \bigcup \left\{L(w) \;\middle|\; w \in \gamma^{-1}(g)\right\}$ and this is a finite union. ◀

Clearly, we have for all $G$: if $K \in \mathrm{SD}_G(A^*)$ is a prefix code of bounded synchronization delay, then $K^*$ and $K^\omega$ are both in $\mathrm{SD}_G(A^\infty)$. As a special case, using the prefix code $K = \emptyset$, it holds $K^* = \{1\} \in \mathrm{SD}_G(A^\infty)$. More generally, every finite language is in $\mathrm{SD}_G(A^\infty)$. Note also that for $G' \le G$ we have $\mathrm{SD}_{G'}(A^\infty) \subseteq \mathrm{SD}_G(A^\infty)$. In particular, $\bigcup \{\mathrm{SD}_{G_i}(A^\infty) \mid i \in I\} \subseteq \mathrm{SD}_{\prod_{i \in I} G_i}(A^\infty)$ for every finite index set $I$. This inclusion holds for every divisor of $G$ as observed by the next lemma which can be proved by induction.

▶ **Lemma 3.** $\mathrm{SD}_H(A^\infty) \subseteq \mathrm{SD}_G(A^\infty)$ *holds for $H \preceq G$.*

We will formulate our some of results on the language classes $\mathrm{SD}_G(A^\infty)$ to obtain finer results. However, our main result concerns the language class

$$\mathrm{SD}_\mathbf{H}(A^\infty) = \bigcup \{\mathrm{SD}_G(A^\infty) \mid G \in \mathbf{H}\}.$$

▶ **Theorem 4.** *Let* $\mathbf{H}$ *be a variety of finite groups. Then* $\overline{\mathbf{H}}(A^\infty)$ *is the smallest class of languages* $\mathcal{C}$ *closed under finite union, concatenation,* $\mathbf{H}$*-controlled star and* $\mathbf{H}$*-controlled* $\omega$*-power such that* $\mathcal{C}$ *contains all finite languages over* $A^*$*. In other words, it holds* $\overline{\mathbf{H}}(A^\infty) = \mathrm{SD}_{\mathbf{H}}(A^\infty)$*.*

▶ **Corollary 5.** $\mathrm{SD}_{\mathbf{H}}(A^\infty)$ *is closed under complementation and intersection for every variety* $\mathbf{H}$ *of finite groups.*

An algebraic characterization of $\overline{\mathbf{H}}$ in terms of Rees extensions will be given in Theorem 15. The proof of Theorem 4 covers the next two sections.

## 3 Closure properties of $\mathrm{SD}_{\mathbf{H}}$

In this section we prove $\mathrm{SD}_{\mathbf{H}}(A^\infty) \subseteq \overline{\mathbf{H}}(A^\infty)$. Therefore one has to study the closure properties under the operations given in the definition of $\mathrm{SD}_{\mathbf{H}}(A^\infty)$, that is, one has to show that those operations do not introduce new groups.

The next proposition shows that the $\mathbf{H}$-controlled star does not introduce new groups.

▶ **Proposition 6** ([16])**.** *Let* $K = \bigcup \{K_g \mid g \in G\} \subseteq A^+$ *be a prefix code of bounded synchronization delay where each* $K_g$ *is regular. Then all subgroups in the syntactic monoid of the* $G$*-controlled star are divisors either of* $G$ *or of the direct product* $\prod_{g\in G} \mathrm{Synt}(K_g)$*.*

We will prove the same for $\gamma^{-1}(1)^\omega$, relying on Proposition 6 as a blackbox. The concept used for transfering the properties to infinite words are Birget-Rhodes expansions [3, 4]. The Birget-Rhodes expansion of a monoid $M$ is the monoid $\mathrm{Exp(M)} = \{(X, m) \mid 1, m \in X \subseteq M\}$. The multiplication on Exp(M) is given as a semi-direct product: $(X, m)\cdot(Y, n) = (X \cup m\cdot Y, m\cdot n)$. Note that $M$ is isomorphic to the submonoid $\{(M, m) \mid m \in M\}$ of Exp(M), that is, $M$ is a divisor of Exp(M). Moreover, the following lemma shows that the Birget-Rhodes expansion has the same groups as $M$.

▶ **Lemma 7.** *Every subgroup of* $\mathrm{Exp(M)}$ *is isomorphic to some group in* $M$*.*

**Proof.** Let $G \subseteq \mathrm{Exp(M)}$ be a group contained in Exp(M) and let $(X, e) \in G$ be the unit in $G$. For every element $(Y, m) \in G$ we have $(X, e)(Y, m) = (X \cup eY, em) = (Y, m)$ and hence, $X \subseteq Y$. Furthermore, $(Y, m)^{|G|} = (Y \cup \cdots, m^{|G|}) = (X, e)$ and we conclude $X = Y$. Thus, $(X, m) \mapsto m$ is an injective embedding of $G$ into $M$. ◀

The idea behind the Birget-Rhodes expansion is that it stores the seen prefixes in a set.

▶ **Lemma 8.** *Let* $\varphi : A^* \to M$ *be a homomorphism and* $\psi : A^* \to \mathrm{Exp(M)}$ *be the homomorphism given by* $\psi(a) = (\{1, \varphi(a)\}, \varphi(a))$*. Let* $u \in A^*$ *and* $\psi(u) = (X, \varphi(u))$*. For every* $m \in X$ *there exists a prefix* $v$ *of* $u$ *such that* $\varphi(v) = m$*.*

**Proof.** We will prove this inductively. The statement is true if $u$ is the empty word. Thus, consider $u = va$ for some letter $a \in A$. Let $\psi(v) = (Y, \varphi(v))$, then

$$\psi(u) = \psi(v) \cdot (\{1, \varphi(a)\}, \varphi(a)) = (Y \cup \{\varphi(v), \varphi(v)\varphi(a)\}, \varphi(u)).$$

Inductively, we obtain prefixes of $v$, and therefore also prefixes of $u$, for all elements of $Y$. The only (potentially) new element in $X$ is $\varphi(u)$. This proves the claim. ◀

A special kind of $\omega$-regular languages are *arrow languages*. Let $L \subseteq A^*$ be a language. We define $\overrightarrow{L} = \{\alpha \in A^\omega \mid \text{infinitely many prefixes of } \alpha \text{ are in } L\}$ to be the arrow language of $L$. The set of arrow languages is exactly the set of deterministic languages [21]. The Birget-Rhodes expansion can be used to obtain a recognizing monoid for $\overrightarrow{L}$, given a monoid for $L$. For a related result see [10].

▶ **Proposition 9.** *Let $L \subseteq A^*$ be some regular language and $\varphi : A^* \to M$ be a homomorphism which recognizes $L$, then $\overrightarrow{L}$ is recognized by* $\mathrm{Exp(M)}$.

**Proof.** Let $\psi : A^* \to \mathrm{Exp(M)}$ be the homomorphism given by $\psi(a) = (\{1, \varphi(a)\}, \varphi(a))$. Let $\alpha \in \overrightarrow{L}$ and $\alpha \sim_\psi \beta$. We show that $\beta \in \overrightarrow{L}$. Let $\alpha = u_1 u_2 \cdots$ and $\beta = v_1 v_2 \cdots$ be factorizations such that $\psi(u_i) = \psi(v_i)$. Since $\alpha \in \overrightarrow{L}$, we may assume that for every $i$ there exists a decomposition $u_i = u_i' u_i''$ such that $u_1 \cdots u_{i-1} u_i' \in L$. By $\psi(u_i) = \psi(v_i)$ and Lemma 8, there exists a decomposition $v_i = v_i' v_i''$ such that $\varphi(u_i') = \varphi(v_i')$. Thus, $u_1 \cdots u_{i-1} u_i' \sim_\varphi v_1 \cdots v_{i-1} v_i'$ and therefore $v_1 \cdots v_{i-1} v_i' \in L$. This implies $\beta \in \overrightarrow{L}$. ◀

▶ **Proposition 10.** *If $L \in \mathrm{SD}_G(A^\infty)$, then all subgroups in* $\mathrm{Synt}(L)$ *are a divisor of a direct product of copies of $G$.*

**Proof.** We will prove this inductively on the definition of $\mathrm{SD}_G(A^\infty)$. The cases $\emptyset \in \mathrm{SD}_G(A^\infty)$ and $\{a\} \in \mathrm{SD}_G(A^\infty)$ for all letters $a \in A$ are straightforward, as they are recognized by aperiodic monoids. Let $L, K$ be languages, such that their syntactic monoids contain only groups which are divisors of a direct product of $G$. The language $L \cup K$ is recognized by the direct product of their syntactic monoids which implies the statement. $(L \cap A^*) \cdot K$ is recognized by the Schützenberger product of their syntactic homomorphisms [10] and [8, Proposition 11.7.10]. The Schützenberger product does not introduce new groups [15].

Let $K \subseteq A^+$ be a prefix code of bounded synchronization delay and $\gamma : K^* \to G$ be a homomorphism of the free monoid $K^*$ to the group $G$ such that for all $g \in G$ every subgroup of $\mathrm{Synt}(K \cap \gamma^{-1}(g))$ is a divisor of a direct product of copies of $G$. Proposition 6 implies that every subgroup of $\mathrm{Synt}(\gamma^{-1}(1))$ is a divisor of a direct product of copies of $G$. Note that $\gamma^{-1}(1)^\omega = \overrightarrow{\gamma^{-1}(1)}$ and therefore Proposition 9 and Lemma 7 imply that every subgroup of $\mathrm{Synt}(\gamma^{-1}(1)^\omega)$ is a divisor of a direct product of copies of $G$. ◀

## 4 The inclusion $\overline{\mathbf{H}}(A^\infty) \subseteq \mathbf{SD}_{\mathbf{H}}(A^\infty)$

We prove that if every subgroup of $M$ is a divisor of $G$, then every language recognized by $M$ is contained in $\mathrm{SD}_G(A^\infty)$. This result is again finer than just the inequality $\overline{\mathbf{H}}(A^\infty) \subseteq \mathrm{SD}_{\mathbf{H}}(A^\infty)$. The proof works by induction on $|M|$ and on the alphabet and decomposes every $\approx_\varphi$-class into several sets in $\mathrm{SD}_G(A^\infty)$. As a byproduct we obtain a normal form for the languages in $\mathrm{SD}_G(A^\infty)$.

▶ **Proposition 11.** *Let $L \subseteq A^\infty$ be recognized by $\varphi : A^* \to M$ and let $G$ be a group such that every subgroup of $M$ is a divisor of $G$, then $L \in \mathrm{SD}_G(A^\infty)$. Moreover, $L$ can be written as finite union*

$$L = L_0 \cup \bigcup_{i=1}^{m} L_i \cdot \gamma_i^{-1}(1)^\omega$$

*for $L_i \in \mathrm{SD}_G(A^*)$ and $\gamma_i : K_i^* \to G$ for prefix codes $K_i \in \mathrm{SD}_G(A^*)$ of bounded synchronization delay with $\gamma_i^{-1}(g) \cap K_i \in \mathrm{SD}_G(A^*)$ for all $g \in G$. All products in the expressions of $L_i$ are unambiguous.*

**Proof.** Let $[\![w]\!]_\varphi = \{v \in A^\infty \mid w \approx_\varphi v\}$ be the equivalence class of $w$. Since $L$ is recognized by $\varphi$, it holds $L = \cup_{w \in L} [\![w]\!]_\varphi$. Our goal is to construct languages $L(w) \in \mathrm{SD}_G(A^\infty)$ such that

- $w \in L(w) \subseteq [\![w]\!]_\varphi$.
- the number of such languages is finite.
- every word in $L(w)$ starts with the same letter.

In particular, we want to saturate $[\![w]\!]_\varphi$ by sets in $\mathrm{SD}_G(A^\infty)$. The construction of the set $L(w)$ is by induction on $(|M|, |A|)$ with lexicographic order.

If $w = 1$, then we set $L(w) = \{1\}$. This concludes the induction base $|A| = 0$. Let us consider the case that $\varphi(A^*)$ is a group, that is, a divisor of $G$. Let $K = A$. The set $K$ is a prefix code of synchronization delay 0 and we may choose the homomorphism $\gamma = \varphi$. Note that every subset of $A$ is in $\mathrm{SD}_G(A^*)$. In particular, $K_g = K \cap \gamma^{-1}(g) \in \mathrm{SD}_G(A^*)$ for all $g \in \varphi(A^*)$. This shows $\gamma^{-1}(g) = \varphi^{-1}(g) \in \mathrm{SD}_G(A^*)$ for all $g \in \varphi(A^*)$ by Lemma 2 and Lemma 3. In order to satisfy the third condition let $w = av \in aA^*$ for some $a \in A$ and set $L(w) = a\varphi^{-1}(\varphi(v))$. It is clear that $w \in L(w) \subseteq [\![w]\!]_\varphi$ and $L(w) \in \mathrm{SD}_G(A^*)$ by the above. If $w \in aA^\omega$, then we obtain $w \in a\varphi^{-1}(g)\varphi^{-1}(1)^\omega$ for some $g \in \varphi(A^*)$ by the pigeonhole principle. Thus, we may set $L(w) = a\varphi^{-1}(g)\varphi^{-1}(1)^\omega$. Note that by the definition of $\sim_\varphi$, the inclusion $L(w) \subseteq [\![w]\!]_\varphi$ holds. In particular, these cases include the induction base $|M| = 1$.

In the following we assume that $\varphi(A^*)$ is not a group and therefore there exists a letter $c \in A$ such that $\varphi(c)$ is not a unit. Fix this letter $c \in A$ and set $B = A \setminus \{c\}$. If $w \in B^\infty$, the set $L(w)$ exists by induction. Let $w = uv$ with $u \in B^*$ and $v \in cA^\infty$. By induction we obtain $L(u) \in \mathrm{SD}_G(B^\infty) \subseteq \mathrm{SD}_G(A^\infty)$ and it remains to show $L(v) \in \mathrm{SD}_G(A^\infty)$. Note that the product $L(w) = L(u) \cdot L(v)$ is unambiguous. From now on we may assume $w \in cA^\infty$. Let us first consider the case $w = uv$ with $u \in c(B^*c)^*$ and $v \in B^\infty$, i.e., there are only finitely many occurrences of the letter $c$ in $w$. By induction, there exists $L(v) \in \mathrm{SD}_G(B^\infty) \subseteq \mathrm{SD}_G(A^\infty)$ and by setting $L(w) = L(u) \cdot L(v)$ it remains to construct $L(u)$.

Consider the alphabet $T = \varphi(B^*) = \{\varphi(u) \mid u \in B^*\}$. Let $M_c$ be the local divisor of $M$ at $\varphi(c)$. Since $M_c$ is a divisor of $M$, every subgroup of $M_c$ is a divisor of $G$. Consider the homomorphism $\psi : T^* \to M_c$ given by $\psi(\varphi(u)) = \varphi(cuc)$ and the substitution $\sigma : (B^*c)^\infty \to T^\infty$ with $\sigma(u_1cu_2c\ldots) = \varphi(u_1)\varphi(u_2)\cdots$. Note that

$$\begin{aligned}\psi(\sigma(u_1cu_2c\ldots u_nc)) &= \psi(\varphi(u_1)\varphi(u_2)\cdots\varphi(u_n)) = \varphi(cu_1c) \circ \varphi(cu_2c) \circ \cdots \circ \varphi(cu_nc)\\ &= \varphi(cu_1cu_2c\ldots cu_nc)\end{aligned}$$

and thus $\varphi^{-1}(m) \cap c(B^*c)^* = c\sigma^{-1}(\psi^{-1}(m))$. Since $|M_c| < |M|$, we can apply induction on the monoid size and there exists a language $L(\sigma(u')) \in \mathrm{SD}_G(T^\infty)$ for all $u' \in (B^*c)^*$. We set $L(u) = c\sigma^{-1}(L(\sigma(u')))$ for $u = cu'$. In order to complete the case of finitely many $c$'s, it suffices to show the following claim:

▶ **Claim.** *It is $\sigma^{-1}(K) \in \mathrm{SD}_G(A^\infty)$ for all $K \in \mathrm{SD}_G(T^\infty)$.*

**Proof of the Claim:** We prove the claim inductively on the definition of $\mathrm{SD}_G$. For $K = \emptyset$, we obtain $\sigma^{-1}(K) = \emptyset \in \mathrm{SD}_G(A^\infty)$. Furthermore,

$$\sigma^{-1}(t) = \bigcup_{v \in B^*, t = \varphi(v)} L(v)c \in \mathrm{SD}_G(A^\infty).$$

Let $L, K \in \mathrm{SD}_G(T^\infty)$. A basic result from set theory yields $\sigma^{-1}(L \cup K) = \sigma^{-1}(L) \cup \sigma^{-1}(K)$. Let $\sigma(v) = w_1w_2$ for some $v \in (B^*c)^*$. Since $B^*c$ is a prefix code, there exists a unique factorization $v = v_1v_2$ with $v_1, v_2 \in (B^*c)^*$ such that $\sigma(v_1) = w_1$ and $\sigma(v_2) = w_2$.

Thus, we conclude $\sigma^{-1}(K \cdot L) = \sigma^{-1}(K) \cdot \sigma^{-1}(L)$. Let now $K \in \mathrm{SD}_G(T^\infty)$ be a prefix code of synchronization delay $d$. We first show that $\sigma^{-1}(K)$ is a prefix code of bounded synchronization delay. Let $u, uv \in \sigma^{-1}(K)$, then $\sigma(u), \sigma(uv) = \sigma(u)\sigma(v) \in K$ and therefore $\sigma(v) = 1$. This implies $v = 1$ and $\sigma^{-1}(K)$ is a prefix code. We prove that $\sigma^{-1}(K)$ has synchronization delay $d+1$. The incrementation of the synchronization delay by one comes from the fact that $B^*c$ is not a suffix code, and thus we need another word in $B^*c$ to pose as a left marker. Consider $uvw \in \sigma^{-1}(K)^*$ with $v \in \sigma^{-1}(K)^{d+1}$ and factorize $v = v_1 c v_2$ with $v_2 \in \sigma^{-1}(K)^d = \sigma^{-1}(K^d)$. Then $\sigma(uvw) = \sigma(uv_1c)\sigma(v_2)\sigma(w)$, and by $\sigma(v_2) \in K^d$ this implies $\sigma(uv) = \sigma(uv_1c)\sigma(v_2) \in K^*$. Thus, $uv \in \sigma^{-1}(K)^*$. Let $\gamma : K^* \to G$ be some homomorphism and $K_g = K \cap \gamma^{-1}(g) \in \mathrm{SD}_G(T^\infty)$ for all $g \in G$. Inductively, $\sigma^{-1}(K_g) \in \mathrm{SD}_G(A^\infty)$ and $\sigma^{-1}(K) = \bigcup \sigma^{-1}(K_g)$. Let $\gamma' : \sigma^{-1}(K)^* \to G$ be induced by $\gamma'(u) = \gamma(\sigma(u))$. By definition of $\mathrm{SD}_G(A^\infty)$ we obtain $\gamma'^{-1}(1) \in \mathrm{SD}_G(A^\infty)$. However, $u_1 \cdots u_n \in \sigma^{-1}(\gamma^{-1}(1))$ if and only if $\gamma(\sigma(u_1 \cdots u_n)) = 1$. Furthermore, note that $\gamma(\sigma(u_1 \cdots u_n)) = \gamma(\sigma(u_1)) \cdots \gamma(\sigma(u_n)) = \gamma'(u_1) \cdots \gamma'(u_n) = \gamma'(u_1 \cdots u_n)$. Thus, we obtain $\sigma^{-1}(\gamma^{-1}(1)) = \gamma'^{-1}(1) \in \mathrm{SD}_G(A^\infty)$ and $\sigma^{-1}(\gamma^{-1}(1)^\omega) = \gamma'^{-1}(1)^\omega \in \mathrm{SD}_G(A^\infty)$. This concludes the proof of the claim. ◀

At this point we showed the proposition for languages $L \subseteq A^*$.

The last case of the proof is that $w$ contains infinitely many $c$'s, that is, $w = cv$ with $v \in (B^*c)^\omega$. By induction, we know that $\sigma(v) \in L_T \cdot \gamma_T^{-1}(1)^\omega \subseteq [\![\sigma(v)]\!]_\psi$ for some $L_T \in \mathrm{SD}_G(T^*)$ and $\gamma_T : K_T^* \to G$ for some prefix code $K_T \in \mathrm{SD}_G(T^*)$ of bounded synchronization delay with $\gamma_T^{-1}(g) \cap K_T \in \mathrm{SD}_G(T^*)$. By the calculation above, there exists a $\gamma : K^* \to G$ with the usual properties such that $\gamma^{-1}(1) = \sigma^{-1}(\gamma_T^{-1}(1))$. Let $L = \sigma^{-1}(L_T)$ and set $L(w) = cL\gamma^{-1}(1)^\omega$. It remains to show that $cL\gamma^{-1}(1)^\omega \subseteq [\![w]\!]_\varphi$. Let $cu \in cL\gamma^{-1}(1)^\omega$, then $\sigma(u) \in [\![\sigma(v)]\!]_\psi$, that is $\sigma(u) \approx_\psi \sigma(v)$. Since $\approx_\psi$ is the transitive closure of $\sim_\psi$, we show that $\sigma(u) \sim_\psi \sigma(v)$ implies $cu \approx_\varphi cv$ for all $u, v \in (B^*c)^\omega$ which concludes the proof. Now, let $\sigma(u) = \sigma(u_1c)\sigma(u_2c)\cdots$ and $\sigma(v) = \sigma(v_1c)\sigma(v_2c)\cdots$ such that $\psi(\sigma(u_ic)) = \psi(\sigma(v_ic))$. As observed above, this implies $\varphi(cu_ic) = \varphi(cv_ic)$. Thus,

$$\begin{aligned}
cu &= (cu_1c)u_2(cu_3c)u_4(c\cdots \sim_\varphi (cv_1c)u_2(cv_3c)u_4(c\cdots \\
&= cv_1(cu_2c)v_3(cu_4c)\cdots \sim_\varphi cv_1(cv_2c)v_3(cv_4c)\cdots \\
&= cv.
\end{aligned}$$

This implies the existence of finitely many sets $L(w) \in \mathrm{SD}_G(A^\infty)$ with $w \in L(w) \subseteq [\![w]\!]_\varphi$ in the case of infinitely many $c$'s. ◀

## 5 Rees extension monoids

We need the fact that every group contained in $\mathrm{Rees}(\mathrm{N}, \mathrm{M}, \rho)$ is contained in $N$ or in $M$.

▶ **Lemma 12** ([1]). *Let $G$ be a subgroup of $\mathrm{Rees}(\mathrm{N}, \mathrm{M}, \rho)$, then there exists an embedding of $G$ into $N$ or into $M$.*

Thus, Lemma 12 implies $\mathrm{LocRees}(\mathbf{H}) \subseteq \mathrm{Rees}(\mathbf{H}) \subseteq \mathrm{Rees}(\overline{\mathbf{H}}) \subseteq \overline{\mathbf{H}}$ for any group variety $\mathbf{H}$. We want to prove equality, that is, every monoid which contains only groups in $\mathbf{H}$ is a divisor of an iterated Rees extension of groups in $\mathbf{H}$. However, we are able to prove a stronger statement using only local Rees extensions.

▶ **Lemma 13.** *Let $M$ be a monoid, $N$ be a submonoid of $M$ and $c \in M$. If $N$ and $c$ generate $M$, then $M$ is a homomorphic image of the local Rees extension $\mathrm{LocRees}(\mathrm{N}, \mathrm{M_c})$.*

**Proof.** Let $\varphi : \mathrm{LocRees(N, M_c)} \to M$ be the mapping given by $\varphi(n) = n$ for $n \in N$ and $\varphi(u, x, v) = uxv$ for $(u, x, v) \in N \times M_c \times N$. Since

$$\begin{aligned}\varphi((u, x, v)(s, y, t)) &= \varphi(u, x \circ cvsc \circ y, t) = \varphi(u, xvsy, t)\\ &= (uxv)(syt) = \varphi(u, x, v)\varphi(s, y, t),\end{aligned}$$

$\varphi$ is a homomorphism. Obviously, $M = N \cup NM_cN$ and thus $\varphi$ is surjective. ◀

A *Rees decomposition* of a monoid $M$ is a sequence of monoids $M_1, \ldots, M_k = M$ such that for each $1 \leq j \leq k$ we have for $M_j$ one of the following:

- $M_j$ is a group which is a divisor of $M$.
- $M_j$ is a divisor of a local Rees extension of a submonoid $M_i$ of $M_j$ and a local divisor $M_\ell$ of $M_j$ with $i, \ell < j$.

▶ **Proposition 14.** *A finite monoid $M$ has a Rees decomposition of length at most $2^{|M|} - 1$.*

**Proof.** We prove the statement with induction on $|M|$. If $M$ is a group, we set $M_1 = M$. This includes the base case $|M| = 1$. If $M$ is not a group, we may choose a minimal generating set of $M$. Let $c$ be a nonunit of this generating set, then there exists a proper submonoid $N$ of $M$ such that $N$ and $c$ generate $M$. Since $c$ is not a unit, the local divisor $M_c$ is smaller than $M$, that is, $|M_c| < |M|$. By induction, there exist Rees decompositions $M'_1, \ldots, M'_{k'} = N$ and $M''_1, \ldots, M''_{k''} = M_c$ with $k', k'' \leq 2^{|M|-1} - 1$. Note that every group, which is a divisor of $N$ or $M_c$ also is a divisor of $M$. Furthermore, $M$ is a divisor of the local Rees extension of $M_{k'} = N$ and $M_{k'+k''} = M_c$ by Lemma 13. Therefore, choosing

- $M_i = M'_i$ for $1 \leq i \leq k'$
- $M_{i+k'} = M''_i$ for $1 \leq i \leq k''$
- $M_{k'+k''+1} = M$

leads to such a sequence for $M$. Since $k' + k'' + 1 \leq 2 \cdot (2^{|M|-1} - 1) + 1 = 2^{|M|} - 1$, the bound on $k$ holds. ◀

The inclusion $\overline{\overline{\mathbf{H}}} \subseteq \mathrm{LocRees}(\mathbf{H})$ is immediate from Proposition 14. This yields

▶ **Theorem 15.** *Let $\mathbf{H}$ be a variety of finite groups. Then $\overline{\overline{\mathbf{H}}} = \mathrm{LocRees}(\mathbf{H}) = \mathrm{Rees}(\mathbf{H})$.*

In particular, every monoid in $\overline{\overline{\mathbf{H}}}$ is a divisor of an iterated Rees extension of groups in $\mathbf{H}$ by Lemma 1. We can draw the decomposition as a tree based on the decomposition of $M$ in submonoids and local divisors. We do not describe this formally but content ourselves to give an example.

▶ **Example 16.** Let $M$ be the monoid generated by $\{a, b, \delta, \sigma\}$ with the relations $a^2 = b^2 = ab = ba = 0$, $a\delta = a$, $\delta\sigma = \sigma\delta^2$, $\delta^3 = 1$, $\sigma^2 = 1$ and $d\delta = \delta d$, $d\sigma = \sigma d$ with $d \in \{a, b\}$. The subgroup generated by $\delta$ and $\sigma$ is the symmetric group $\mathfrak{S}_3$; it is solvable but not abelian. The monoid $M$ is syntactic for the language $L$ which is a union of $L_a$ and $L_b$. The language $L_a$ is the set of all words $uav$ with $uv \in \{\delta, \sigma\}^*$ and the sign of the permutation $uv$ evaluates to $-1$. The language $L_b$ is the set of all words $ubv$ with $uv \in \{\delta, \sigma\}^*$ and $uv$ evaluates in $\mathfrak{S}_3$ to $\delta$. The decomposition in local Rees extensions from Proposition 14 is depicted in Figure 1. Here $M[a, \sigma, \delta]$ denotes the submonoid generated by $\{a, \sigma, \delta\}$. In particular, this yields

$$M \preceq \mathrm{Rees}(\mathrm{Rees}(\mathfrak{S}_3, \mathrm{Rees}(\mathbb{Z}/2\mathbb{Z}, \{1\}, \rho_1), \rho_2), \mathrm{Rees}(\mathfrak{S}_3, \{1\}, \rho_3), \rho_4)$$

for some $\rho_1, \rho_2, \rho_3, \rho_4$ by Lemma 1.

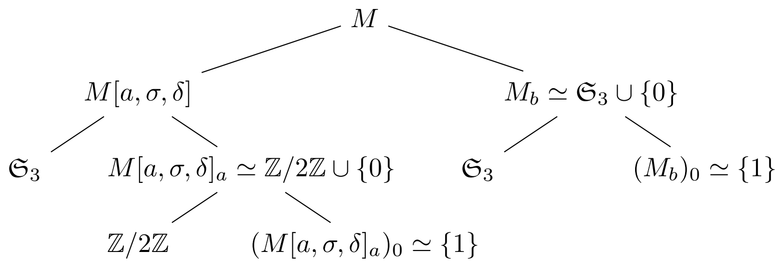


**Figure 1** Decomposition tree of the monoid in Example 16.

## 6 Applications

An application of Proposition 14 is the solution to an open problem of Almeida and Klíma. Let $\mathbf{U}$ and $\mathbf{V}$ be varieties. Let $\mathrm{Rees}(\mathbf{U}, \mathbf{V})$ be the variety generated by $\mathrm{Rees}(\mathrm{N}, \mathrm{M}, \rho)$ for $N \in \mathbf{U}$ and $M \in \mathbf{V}$. Note that in general $\mathrm{Rees}(\mathbf{V}) \neq \mathrm{Rees}(\mathbf{V}, \mathbf{V})$. However $\mathrm{Rees}(\mathbf{V})$ can be defined as the limit of this operation. Let $\mathbf{V}_i = \mathrm{Rees}(\mathbf{V}_{\mathrm{i}-1}, \mathbf{V}_{\mathrm{i}-1})$ and $\mathbf{V}_0 = \mathbf{V}$, then

$$\mathrm{Rees}(\mathbf{V}) = \bigcup_{i \in \mathbb{N}} \mathbf{V}_i.$$

The variety $\mathrm{Rees}(\mathbf{U}, \mathbf{V})$ has recently been introduced by Almeida and Klíma under the name of *bullet operation* [1]. They defined a variety $\mathbf{V}$ to be *bullet idempotent* if $\mathbf{V} = \mathrm{Rees}(\mathbf{V}, \mathbf{V})$ and they asked whether there are varieties apart from $\overline{\mathbf{H}}$ which are bullet idempotent. Using our decomposition above, we prove that the answer to this question is "No".

▶ **Theorem 17.** *Let* $\mathbf{V}$ *be a bullet idempotent variety and let* $\mathbf{H} = \mathbf{V} \cap \mathbf{G}$*, then* $\mathbf{V} = \overline{\mathbf{H}}$*.*

**Proof.** Since $\overline{\mathbf{H}}$ is the maximal variety with $\overline{\mathbf{H}} \cap \mathbf{G} = \mathbf{H}$, we have $\mathbf{V} \subseteq \overline{\mathbf{H}}$. Let $M \in \overline{\mathbf{H}}$. Inductively, we may assume that every proper divisor of $M$ is in $\mathbf{V}$. If $M$ is a group, then $M \in \mathbf{H}$ and thus $M \in \mathbf{V}$. Thus, there exists an nonunit element $c \in M$ and a proper submonoid $N$ of $M$ such that $N$ and $c$ generate $M$. By Lemma 13, $M$ is a divisor of $\mathrm{LocRees}(\mathrm{N}, \mathrm{M_c})$, and since $N, M_c \in \mathbf{V}$ and $\mathbf{V} = \mathrm{Rees}(\mathbf{V}, \mathbf{V})$ we obtain $M \in \mathbf{V}$. ◀

Let $(\mathrm{FO} + \mathrm{MOD}_q)[<]$ be the fragment of first-order sentences which only use first-order quantifiers, modular quantifiers of modulus $q$ and the predicate $<$. Then the following theorem holds.

▶ **Corollary 18.** $(\mathrm{FO} + \mathrm{MOD}_q)[<](A^\infty) = \mathrm{SD}_{\mathbf{Sol}_q}(A^\infty)$

**Proof.** By [20], see also [19] for a complete treatise, $(\mathrm{FO} + \mathrm{MOD}_q)[<]$ describes the family of all regular languages such that every group in the syntactic monoid is a solvable group of cardinality dividing a power of $q$, that is the languages in $\mathbf{Sol}_q$. Theorem 4 then implies the stated equality. ◀

The same language class has been described by Straubing with another operation, counting how many prefixes are in a given language, which resembles more closely the counting modulo $q$ [18].

## 7 Summary

Our main theorem Theorem 4 states $\overline{\mathbf{H}}(A^\infty) = \mathrm{SD}_{\mathbf{H}}(A^\infty)$. An overview over the contributions for $\overline{\mathbf{H}}$ is given in Table 1.

**Table 1** Overview of existing and new language characterizations of $\overline{\mathbf{H}}$.

| | $\overline{\mathbf{1}}$ | $\overline{\mathbf{Ab}}$ | $\overline{\mathbf{Sol}}$ | $\overline{\mathbf{Sol}_q}$ | $\overline{\mathbf{H}}$ |
|---|---|---|---|---|---|
| finite words | [17] | [16] | [18],**new** | [18],**new** | **new, unless** $\mathbf{H} \subseteq \mathbf{Ab}$ |
| $\omega$-words | [5] | **new** | **new** | **new** | **new, unless** $\mathbf{H} = \mathbf{1}$ |

As a byproduct we were able to give a simple decomposition of the monoids in $\overline{\mathbf{H}}$ as local Rees extensions and groups in $\mathbf{H}$, using only exponentially many operations.

## References


1. Jorge Almeida and Ondřej Klíma. On the irreducibility of pseudovarieties of semigroups. *Journal of Pure and Applied Algebra*, 220(4):1517–1524, 2016. doi:10.1016/j.jpaa.2015.09.015.
2. André Arnold. A syntactic congruence for rational $\omega$-languages. *Theoretical Computer Science*, 39:333–335, 1985.
3. Jean-Camille Birget and John L. Rhodes. Almost finite expansions of arbitrary semigroups. *Journal of Pure and Applied Algebra*, 32(3):239–287, 1984.
4. Jean-Camille Birget and John L. Rhodes. Group theory via global semigroup theory. *Journal of Algebra*, 120(2):284–300, 1989. doi:10.1016/0021-8693(89)90199-3.
5. Volker Diekert and Manfred Kufleitner. Omega-rational expressions with bounded synchronization delay. *Theory Comput. Syst.*, 56:686–696, 2015.
6. Volker Diekert and Manfred Kufleitner. A survey on the local divisor technique. *Theoretical Computer Science*, 610:13–23, 2015. doi:10.1016/j.tcs.2015.07.008.
7. Volker Diekert, Manfred Kufleitner, and Pascal Weil. Star-free languages are Church-Rosser congruential. *Theoretical Computer Science*, 454:129–135, 2012. doi:10.1016/j.tcs.2012.01.028.
8. Volker Diekert and Grzegorz Rozenberg, editors. *The Book of Traces*. World Scientific, Singapore, 1995.
9. Samuel Eilenberg. *Automata, Languages, and Machines*, volume B. Academic Press, New York and London, 1976.
10. Dominique Perrin. Recent results on automata and infinite words. In *Mathematical foundations of computer science, 1984 (Prague, 1984)*, volume 176 of *Lecture Notes in Comput. Sci.*, pages 134–148. Springer, Berlin, 1984.
11. Dominique Perrin and Jean-Éric Pin. *Infinite words*, volume 141 of *Pure and Applied Mathematics*. Elsevier, Amsterdam, 2004.
12. Jean-Éric Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.
13. John Rhodes and Dennis Allen. Synthesis of the classical and modern theory of finite semigroups. *Advances in Mathematics*, 11(2):238–266, 1973. doi:10.1016/0001-8708(73)90010-8.
14. John L. Rhodes and Benjamin Steinberg. *The* q*-theory of finite semigroups*. Springer Monographs in Mathematics. Springer, 2009.
15. Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
16. Marcel-Paul Schützenberger. Sur les monoides finis dont les groupes sont commutatifs. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 8(R-1):55–61, 1974.
17. Marcel-Paul Schützenberger. Sur certaines opérations de fermeture dans les langages rationnels. In *Symposia Mathematica, Vol. XV (Convegno di Informatica Teorica, INDAM, Roma, 1973)*, pages 245–253. Academic Press, 1975.

**18** Howard Straubing. Families of recognizable sets corresponding to certain varieties of finite monoids. *Journal of Pure and Applied Algebra*, 15(3):305–318, 1979.

**19** Howard Straubing. *Finite Automata, Formal Logic, and Circuit Complexity.* Birkhäuser, Boston, Basel and Berlin, 1994.

**20** Howard Straubing, Denis Thérien, and Wolfgang Thomas. Regular languages defined with generalized quantifiers. *Inform. and Comput.*, 118(2):289–301, 1995.

**21** Wolfgang Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, chapter 4, pages 133–191. Elsevier Science Publishers B. V., 1990.

**22** Thomas Wilke. An algebraic theory for regular languages of finite and infinite words. *International Journal of Algebra and Computation*, 3(4):447–489, 1993.