

Noname manuscript No. (will be inserted by the editor)
--

Omega-rational expressions with bounded synchronization delay

Volker Diekert · Manfred Kufleitner

the date of receipt and acceptance should be inserted later

Abstract In 1965 Schützenberger published his famous result that star-free languages (SF) and aperiodic languages (AP) coincide over finite words, often written as $SF = AP$. Perrin generalized $SF = AP$ to infinite words in the mid 1980s. In 1973 Schützenberger presented another (and less known) characterization of aperiodic languages in terms of rational expressions where the use of the star operation is restricted to prefix codes with bounded synchronization delay and no complementation is used. We denote this class of languages by SD. In this paper, we present a generalization of $SD = AP$ to infinite words. This became possible via a substantial simplification of the proof for the corresponding result for finite words. Moreover, we show that $SD = AP$ can be viewed as more fundamental than $SF = AP$ in the sense that the classical 1965 result of Schützenberger and its 1980s extension to infinite words by Perrin are immediate consequences of $SD = AP$.

Keywords omega-regular language, star-free language, finite monoid, local divisor, bounded synchronization delay

Mathematics Subject Classification (2000) 20M35 · 68Q45 · 68Q70 · 94A45

CR Subject Classification F.4.3

The second author gratefully acknowledges the support by the German Research Foundation (DFG) under grant DI 435/5-1 and the support by ANR 2010 BLAN 0202 FREC.

FMI, University of Stuttgart,
Universitätsstr. 38, 70569 Stuttgart, Germany
E-mail: {diekert, kufleitner}@fmi.uni-stuttgart.de
Tel. +49 711 685 88231
Fax. +49 711 685 88310

1 Introduction

A fundamental and classical result of Schützenberger [14] says that the class of star-free languages $SF(A^*)$ is exactly the same as the class of aperiodic languages $AP(A^*)$. This result was published in 1965 and later extended to infinite words by Perrin [10] in 1984. We simply write $SF = AP$ when referring to this result. The class of star-free languages is very robust: it also coincides with the class $FO[<]$ of languages definable in first-order logic [8]; and this is the same as the class LTL of languages definable in the linear temporal logic [7]. The equivalence of these characterizations have been established first for finite words and then extended to infinite words.

A proof for the equivalence can be conveniently arranged in a cycle. On this cycle the inclusion $AP \subseteq LTL$ becomes the most difficult one. It is done in the survey [3] with the concept of *local divisors* which play a prominent role here, too.

There is another beautiful characterization of $SF(A^*)$ due to Schützenberger [15], which seems to be quite overlooked. It characterizes $SF(A^*)$ without using complementation, but the inductive definition allows the star-operation on languages K (already belonging to the class) if K is a prefix code with bounded synchronization delay. Since synchronization delay is the main feature in this approach, the class is denoted by $SD(A^*)$. The notion of bounded synchronization delay was introduced by Golomb and Gordon [6] and it is an important concept in coding theory. A proof of $SD(A^*) = SF(A^*)$ can be found e.g. in the textbook by Pin and Perrin on infinite words [11]. But although the book is on infinite words, this result is shown for finite words, only.

The extension to infinite words is actually done for the first time in this paper. This generalization became possible through the technique of local divisors, which also simplifies the classical proof on finite words. Our main result Theorem 1 is slightly more precise than simply stating $SD = AP$ for infinite words: The extension of $SD = AP$ to infinite words has a very nice explicit description where the ω -terms are just over star-free prefix codes of bounded synchronization delay. Moreover, we show that $SD = AP$ can be viewed as more fundamental than $SF = AP$ in the sense that the classical 1965 result of Schützenberger and its 1984 extension to infinite words by Perrin are immediate consequences of $SD = AP$.

We see therefore three contributions in this paper: (1) We considerably simplify the classical proof $SD = AP$ by using the algebraic tool of a local divisor. As a byproduct, we show that unambiguous products are sufficient for defining the class SD . (2) We easily extend $SD = AP$ to infinite words by the very same proof technique. (3) We establish that $SF = AP$ is an immediate consequence of $SD = AP$. This last property can also be seen as an advertisement for the class SD .

2 Preliminaries

In the following A means a finite alphabet and A^* (resp. A^ω) denotes the sets of finite (resp. infinite) words over A . We let $A^\infty = A^* \cup A^\omega$. The empty word is denoted by 1 . The set of nonempty finite words is $A^+ = A^* \setminus \{1\}$. The *product* of two languages $L \subseteq A^*$ and $K \subseteq A^\infty$ is $L \cdot K = \{uv \in A^\infty \mid u \in L, v \in K\}$; the product is *unambiguous* if every word $w \in L \cdot K$ admits a unique factorization $w = uv$ with $u \in L$ and $v \in K$. If $K \subseteq A^+$, then its finite iteration is $K^* = \{v_1 \cdots v_n \mid n \geq 0, v_i \in K\}$ and its infinite iteration is $K^\omega = \{v_1 v_2 \cdots \mid v_i \in K\}$.

A classical result in formal language theory says that a language $L \subseteq A^*$ is regular if and only if it is *recognizable*, see e.g. [12]. The latter means that there is a homomorphism $\varphi : A^* \rightarrow M$ to a finite monoid M such that $L = \varphi^{-1}(\varphi(L))$. If φ is fixed and $u \in A^*$ is a word, we also write $[u]_\varphi$ instead of $\varphi^{-1}(\varphi(u)) = \{v \in A^* \mid \varphi(v) = \varphi(u)\}$. Thus, if $\varphi : A^* \rightarrow M$ recognizes L , then we can write L as

$$L = \bigcup_{u \in L} [u]_\varphi$$

and this union is finite.

A finite monoid M is called *aperiodic* if there exists some positive integer $n \in \mathbb{N}$ such that $x^n = x^{n+1}$ for all $x \in M$. Every submonoid and every homomorphic image of an aperiodic monoid is aperiodic. A language $L \subseteq A^*$ is called *aperiodic* if it is recognized by some finite aperiodic monoid M . The class of aperiodic languages in A^* is denoted by $\text{AP}(A^*)$. Recognizability for ω -words is a little bit more technical to explain than for finite words. We give here a general definition for ∞ -words, which applies to the finitary and to the infinitary case simultaneously, but keeps the ability to distinguish between the empty word, finite non-empty words, and infinite words, respectively.

Every word $u \in A^\infty$ can be written either as a finite or an infinite sequence $u = u_1 u_2 u_3 \cdots$ with $u_i \in A^+$. We have $u_1 u_2 u_3 \cdots \in A^*$ if and only if the sequence is finite, and $u = 1$ if the sequence is empty. The length of the sequence is an element in $\mathbb{N} \cup \{\omega\}$. Given a homomorphism $\varphi : A^* \rightarrow M$, we set

$$u_1 u_2 u_3 \cdots \sim_\varphi v_1 v_2 v_3 \cdots$$

if the two sequences have the same length and if $\varphi(u_i) = \varphi(v_i)$ for all i . A language $L \subseteq A^\infty$ is called *recognizable* or *regular* (resp. *aperiodic*) if there is a homomorphism $\varphi : A^* \rightarrow M$ to a finite (resp. finite and aperiodic) monoid M such that $u \in L$ and $u \sim_\varphi v$ implies $v \in L$ for all $u, v \in A^\infty$. If $L \subseteq A^\infty$ is regular (resp. aperiodic), then $L \cap A^*$ is regular (resp. aperiodic) in the usual sense. Moreover regular (resp. aperiodic) languages of A^* in the sense defined above remain regular (resp. aperiodic) in the new definition. As usual, $\text{AP}(A^\infty)$ denotes the class of all aperiodic languages in A^∞ . Similarly, $\text{AP}(A^\omega)$ contains all aperiodic languages in A^ω . We remark that regular languages in A^ω are sometimes called *omega-regular* or *ω -regular*. We shall use the following simple observation.

Lemma 1 *Let $L \subseteq A^\infty$ be regular. Then the following assertions are equivalent.*

1. $L \in \text{AP}(A^\infty)$.
2. *There exists $n \in \mathbb{N}$ such that for all $u, x, y, z \in A^*$ the following two conditions hold:*

$$\begin{aligned} xu^n yz^\omega \in L &\Leftrightarrow xu^{n+1} yz^\omega \in L \\ x(u^n y)^\omega \in L &\Leftrightarrow x(u^{n+1} y)^\omega \in L. \end{aligned}$$

3. *There exists $n \in \mathbb{N}$ such that for all $u, x, y, z \in A^*$ the following two conditions hold:*

$$\begin{aligned} xu^n yz^\omega \in L &\Rightarrow xu^{n+1} yz^\omega \in L \\ x(u^n y)^\omega \in L &\Rightarrow x(u^{n+1} y)^\omega \in L. \end{aligned}$$

Proof The language L is aperiodic if and only if its syntactic monoid in the sense of Arnold [1] is aperiodic, see e.g. [11]. Hence, 1 is equivalent to 2. The implication from 2 to 3 is trivial. It remains to show that 3 implies 2. It is here where we use the hypothesis that $L \subseteq A^\infty$ is regular. Since L is regular, its syntactic monoid is finite and it recognizes L . Hence there is some $p > 0$ such that, for all n large enough, $xu^n yz^\omega \in L$ if and only if $xu^{n+p} yz^\omega \in L$, and $x(u^n y)^\omega \in L$ if and only if $x(u^{n+p} y)^\omega \in L$. The result follows because by applying the hypothesis $p - 1$ times, we may conclude that $xu^{n+1} yz^\omega \in L$ implies $xu^{n+p} yz^\omega \in L$ and that $x(u^{n+1} y)^\omega \in L$ implies $x(u^{n+p} y)^\omega \in L$. \square

The language $(aa)^*$ is not aperiodic, but $(aa)^\omega = a^\omega$ is aperiodic. However we can infer aperiodicity of K^ω from that of K^* . The following lemma is well-known and it belongs to folklore. For lack of a precise reference and for convenience of the reader we give its proof.

Lemma 2 *If $K^* \in \text{AP}(A^*)$, then $K^\omega \in \text{AP}(A^\infty)$.*

Proof Choose $m \in \mathbb{N}$ such that $xu^m \tilde{y} \in K^*$ implies $xu^{m+1} \tilde{y} \in K^*$ for all $u, x, \tilde{y} \in A^*$. Let $xu^m yz^\omega \in K^\omega$. Then we find a prefix $xu^m \tilde{y} \in K^*$ such that $xu^m yz^\omega = xu^m \tilde{y} \tilde{z}^\omega$. It follows that $xu^n yz^\omega \in K^\omega$ implies $xu^{n+1} yz^\omega \in K^\omega$ for all $n \geq m$. This is the first part in statement 3 of Lemma 1.

Now let $n = 2m$ and consider $x(u^n y)^\omega = v_1 v_2 v_3 \cdots$ with $v_i \in K^+$ for all $i \in \mathbb{N}$. By Lemma 1 it remains to show that $x(u^{n+1} y)^\omega \in K^\omega$. We may assume $u \neq 1$. The infinite sequence $v_1 v_2 v_3 \cdots$ defines cut points in the infinite word $x(u^n y)^\omega$ by choosing the positions between v_i and v_{i+1} . By gathering factors $v_j \cdots v_\ell$ together, we may assume that each factor u^n contains at most one cut point. We can write each $u^n = u^m u^m$ such that either the first u^m or the second u^m is without any cut point, and we can apply the hypothesis at all occurrences of u^n . Thus $x(u^{n+1} y)^\omega \in K^\omega$. \square

3 Local divisors

Let M be a finite monoid and let $c \in M$. Consider the submonoid $M' = \{x \in M \mid xc \in cM\}$. The right-translation by c shifts M' to $M'c = cM \cap Mc$. We turn $cM \cap Mc$ into a monoid by defining a new multiplication \circ on $cM \cap Mc$ by:

$$xc \circ cy = xcy.$$

It is straightforward to see that \circ is well-defined and $(cM \cap Mc, \circ)$ is a monoid with neutral element c . Moreover, $x \mapsto xc$ yields a surjective homomorphism from M' onto $(cM \cap Mc, \circ)$. Thus, $(cM \cap Mc, \circ)$ becomes a divisor of M ; it is called the *local divisor* of M at c . As a consequence, if M is aperiodic, then $(cM \cap Mc, \circ)$ is aperiodic, too. Local divisors were introduced in commutative algebra by Meyberg [9]. In finite semigroup theory they first appear in [2] and have been used, among others, in a recent proof of the Krohn-Rhodes Theorem [5].

Lemma 3 *If M is aperiodic and $c \neq 1$, then $1 \in M \setminus cM$.*

Proof Assume $1 \in cM$. Then there exists $d \in M$ such that $1 = cd$. Since M is aperiodic, we have $c^n = c^{n+1}$ for some $n \in \mathbb{N}$. Now, $1 = c \cdot 1 \cdot d = c^n d^n = c^{n+1} d^n = c$, a contradiction. Thus $1 \in M \setminus cM$. \square

Thus for finite aperiodic monoids M and monoid elements $c \neq 1$ we have $|cM \cap Mc| < |M|$.

4 Schützenberger's class SD

A language $K \subseteq A^*$ is called *prefix-free* if $u, uv \in K$ implies $u = uv$. A prefix-free language $K \subseteq A^+$ is also called a *prefix code* since every word $u \in K^*$ admits a unique factorization $u = u_1 \cdots u_k$ with $k \geq 0$ and $u_i \in K$. 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:

$$\text{if } uvw \in K^* \text{ and } v \in K^d, \text{ then } uv \in K^*.$$

If d is given explicitly, we also say that K has synchronization delay d . Note that every subset $B \subseteq A$ yields a prefix code with synchronization delay 0. In particular, the sets B are prefix codes of bounded synchronization delay for all $B \subseteq A$.

The intuition behind this concept is the following: Assume a sender emits a stream of code words from K , where K is a prefix code with synchronization delay d . If a receiver misses the beginning of the message, he can wait until he detects a sequence of d code words. Then he can synchronize and decipher the remaining text after these d words.

We now inductively define Schützenberger's language class $\text{SD}(A^\infty)$ simultaneously for finite and infinite words as follows:

1. We have $\emptyset \in \text{SD}(A^\infty)$ and $\{a\} \in \text{SD}(A^\infty)$ for all letters $a \in A$.
2. If $L, K \in \text{SD}(A^\infty)$, then $L \cup K \in \text{SD}(A^\infty)$.
3. If $L, K \in \text{SD}(A^\infty)$ and $L \subseteq A^*$, then $L \cdot K \in \text{SD}(A^\infty)$.
4. If $K \in \text{SD}(A^\infty)$ with $K \subseteq A^+$ such that K is a prefix code with bounded synchronization delay, then $K^*, K^\omega \in \text{SD}(A^\infty)$.

Note that (unlike the definition of star-free sets) the inductive definition of $\text{SD}(A^\infty)$ does not use any complementation. The class $\text{SF}(A^\infty)$ of star-free languages is defined similarly:

1. We have $\emptyset \in \text{SF}(A^\infty)$ and $\{a\} \in \text{SF}(A^\infty)$ for all letters $a \in A$.
2. If $L, K \in \text{SF}(A^\infty)$, then $L \cup K \in \text{SF}(A^\infty)$.
3. If $L, K \in \text{SF}(A^\infty)$, then $(L \cap A^*) \cdot K \in \text{SF}(A^\infty)$.
4. If $K \in \text{SF}(A^\infty)$, then $A^\infty \setminus K \in \text{SF}(A^\infty)$.

By $\text{SD}(A^*)$ we denote the finitary subclass of $\text{SD}(A^\infty)$. That is, $\text{SD}(A^*) = \{L \in \text{SD}(A^\infty) \mid L \subseteq A^*\}$. Similarly, $\text{SD}(A^\omega)$ is its infinitary subclass. Also note that $\emptyset^* = \{1\}$ belongs to $\text{SD}(A^*)$ because \emptyset is a prefix code of bounded synchronization delay. The language classes $\text{SF}(A^*)$ and $\text{SF}(A^\omega)$ are defined similarly.

Example 1 Let $B \subseteq A$. For every letter $c \in B$ we let $B_c = B \setminus \{c\}$. Then $B_c^*c \in \text{SD}(A^\infty)$ and B_c^*c is a prefix code with synchronization delay 1. Thus the language $(B_c^*c)^\omega$ of all words in B^ω with infinitely many c is in $\text{SD}(A^\infty)$. \diamond

Lemma 4 *We have $\text{SD}(A^\infty) \subseteq \text{AP}(A^\infty)$.*

Proof The class $\text{AP}(A^\infty)$ contains all finite subsets of A^∞ and it is closed under finite union and concatenation, see e.g. [3]. By Lemma 2 it is enough to show the following claim which concerns only finitary languages: If $K \in \text{SD}(A^*)$ is a prefix code with synchronization delay d , then K^* is aperiodic. By induction we may assume that $K \in \text{AP}(A^*)$. Hence, by Lemma 1, for some $n \in \mathbb{N}$ and all words u, x, y we have $xu^n y \in K$ if and only if $xu^{n+1}y \in K$. Moreover, by the same lemma it is enough to show that $xu^{n(d+1)}y \in K^*$ implies $xu^{n(d+1)+1}y \in K^*$ for all words u, x, y .

Consider $u \in A^+$ and let $m = n(d+1)$. Suppose $xu^m y \in K^+$. There is a unique factorization $xu^m y = v_1 \cdots v_k$ with $v_i \in K$. If u^n occurs as a factor of some v_i , then we are done, because inside such a factor v_i we can replace u^n by u^{n+1} . Therefore, we can assume $xu_1 u_2 \in K^+$ for $u_1 u_2 u_3 = u^{m-1}$ with $u_2 \in K^d$. Since K is a prefix code, we have $q = u_3 u y \in K^*$. The prefix $p = x u u_1 u_2$ of $xu^m y$ ends with a word in K^d . Note the extra u after x in p . Thus, by synchronization delay, we conclude $p \in K^*$. It follows that $xu^{m+1}y = pq \in K^+$. \square

5 SD equals AP

We consider the general situation for languages in A^∞ , i.e., we deal with finite and infinite words simultaneously. Theorem 1 is our main contribution. The

first part of the proof also yields a simple proof for the corresponding result $\text{SD}(A^*) = \text{AP}(A^*)$ over finite words.

Theorem 1 *Let $L \subseteq A^\infty$. The following assertions are equivalent:*

1. $L \in \text{SD}(A^\infty)$.
2. $L \in \text{AP}(A^\infty)$.
3. L can be written as finite union

$$L = L_0 \cup \bigcup_{i=1}^m L_i K_i^\omega,$$

where $L_i, K_i \in \text{SD}(A^*)$, all K_i are prefix codes with bounded synchronization delay, and all products in the rational expressions for L and for each L_i and K_i are unambiguous.

Proof Recall that if $K \in \text{SD}(A^*)$ is a prefix code of bounded synchronization delay, then $K^\omega \in \text{SD}(A^\infty)$ by definition. The implication 3 \Rightarrow 1 is therefore trivial. The implication 1 \Rightarrow 2 is the content of Lemma 4. It remains to show the implication 2 \Rightarrow 3. Note that for $L \subseteq A^*$ this means just to show 2 \Rightarrow 1.

We start with a recognizing homomorphism $\varphi : A^* \rightarrow M$ to a finite aperiodic monoid. We will also show that the languages L_i and K_i appearing in the expression given by 3 can be chosen from a finite collection of sets which only depends on the homomorphism φ .

Let us denote by \approx_φ the equivalence relation on A^∞ generated by \sim_φ , and for each $w \in A^\infty$ let $\llbracket w \rrbracket_\varphi = \{v \in A^\infty \mid v \approx_\varphi w\}$. Since L is regular, there are only finitely many classes $\llbracket w \rrbracket_\varphi$. Note that according to our definition of \sim_φ and \approx_φ we have three possibilities: $\llbracket w \rrbracket_\varphi = \{1\}$ or $\llbracket w \rrbracket_\varphi = [w]_\varphi \setminus \{1\} \subseteq A^+$ or $\llbracket w \rrbracket_\varphi \subseteq A^\omega$.

We show that for every word $w \in A^\infty$ there exists a language $L(w) \in \text{SD}(A^\infty)$ satisfying $w \in L(w) \subseteq \llbracket w \rrbracket_\varphi$ such that the size of the rational expression for $L(w)$ is bounded by a function in $|M|$ and $|A|$; in particular, it does not depend on the length of w . In addition, if w is a finite word, then the expression for $L(w)$ does not use terms of the form K^ω ; and if w is nonempty, then all words in $L(w)$ start with the same letter. It follows $L = \bigcup_{w \in L} L(w)$ and this union is finite since there are only finitely many languages $L(w)$. The construction of $L(w)$ is by induction on the parameter $(|M|, |A|)$ with lexicographic order.

Let $C = \{a \in A \mid a \text{ occurs in } w\}$. First, suppose $|\varphi(C^*)| = 1$. If $w = 1$, we have $L(w) = \{1\}$; if $w \in aA^*$, then we can set $L(w) = aC^*$; and if $w \in A^\omega$, then $L(w) = aC^\omega$. This covers both cases $|M| = 1$ and $A = \emptyset$. Hence, in the remainder of the proof we may assume $\varphi(c) \neq 1$ for some letter $c \in A$ occurring in w . Let $B = A \setminus \{c\}$. We write $w = uv$ with $u \in B^*$ and $v \in cA^\infty$. By induction on the alphabet, there exists $L(u) \in \text{SD}(B^*)$ and it remains to construct $L(v) \in \text{SD}(A^\infty)$. Then $L(w) = L(u)L(v)$, and this product is unambiguous since the occurrence of the first c is unique. In the remainder of the proof we therefore assume $w \in cA^\infty$. We now distinguish two cases: The word w contains only finitely many occurrences of the letter c or it contains infinitely many c .

5.1 Finitely many c

We can write $w = uw'$ with $u \in c(B^*c)^*$ and $w' \in B^\infty$. By induction, there exists $L(w') \in \text{SD}(B^\infty)$ and it suffices to construct $L(u) \in \text{SD}(A^*)$ with $L(u) \subseteq A^*c$. Then we set $L(w) = L(u)L(w')$, and this product is unambiguous. Define a new finite alphabet $T = \{[v]_\varphi \mid v \in B^*\}$ and let $M_c = (\varphi(c)M \cap M\varphi(c), \circ)$ denote the local divisor of M at $\varphi(c)$, see Section 3. Let $\psi : T^* \rightarrow M_c$ be the homomorphism induced by the mapping $[v]_\varphi \mapsto \varphi(cvc)$ for $[v]_\varphi \in T$. Next, we define a mapping (*substitution*) $\sigma : (B^*c)^* \rightarrow T^*$ by $\sigma(v_1c \cdots v_kc) = [v_1]_\varphi \cdots [v_k]_\varphi$ for $v_i \in B^*$. By definition of the operation \circ in M_c we see:

$$\begin{aligned} \psi(\sigma(v_1c \cdots v_kc)) &= \psi([v_1]_\varphi \cdots [v_k]_\varphi) \\ &= \psi([v_1]_\varphi) \circ \cdots \circ \psi([v_k]_\varphi) \\ &= \varphi(cv_1c \cdots v_kc). \end{aligned}$$

Hence, for all $m \in M_c$ we obtain $\varphi^{-1}(m) \cap c(B^*c)^* = c\sigma^{-1}(\psi^{-1}(m))$. Thus, we can set $L(u) = c\sigma^{-1}(L(\sigma(u')))$ for $u = cu'$. By Lemma 3, we have $|M_c| < |M|$ and therefore, by induction on the size of the monoid, $L(\sigma(u')) \in \text{SD}(T^*)$ exists. It remains to show that $K \in \text{SD}(T^\infty)$ implies $\sigma^{-1}(K) \in \text{SD}(A^\infty)$. This last step is done by structural induction over the expression for K . For every $v \in B^*$ there exists a language $L(v) \in \text{SD}(B^*)$ by induction on the size of the alphabet. Hence, for every letter $t \in T$ we have:

$$\sigma^{-1}(t) = \left(\bigcup_{v \in B^*, [v]_\varphi = t} L(v) \right) c$$

and this union is finite. For $K, K' \in \text{SD}(T^*)$ we have:

$$\begin{aligned} \sigma^{-1}(K \cup K') &= \sigma^{-1}(K) \cup \sigma^{-1}(K'), \\ \sigma^{-1}(K \cdot K') &= \sigma^{-1}(K) \cdot \sigma^{-1}(K'), \\ \sigma^{-1}(K^*) &= \sigma^{-1}(K)^*, \\ \sigma^{-1}(K^\omega) &= \sigma^{-1}(K)^\omega. \end{aligned}$$

Note that $\sigma^{-1}(K) \cdot \sigma^{-1}(K')$ is unambiguous provided that $K \cdot K'$ is unambiguous. We still have to verify that $\sigma^{-1}(K)$ is a prefix code of bounded synchronization delay, if K has this property. Clearly, $1 \notin \sigma^{-1}(K)$. To see prefix-freeness, consider $u, uv \in \sigma^{-1}(K)$. This implies $u \in A^*c$ and hence, $\sigma(uv) = \sigma(u)\sigma(v)$. It follows that $v = 1$ because K is prefix-free. Finally, let $L = \sigma^{-1}(K)$ and suppose K has synchronization delay d . We show that L has synchronization delay $d + 1$: Let $uvw \in L^*$ with $v \in L^{d+1}$. Write $v = u'cv'$ with $v' \in L^d$. Note that $v' \in A^*c$. It follows $\sigma(uv) = \sigma(uu'c)\sigma(v')$ and $\sigma(v') \in K^d$. Thus, $\sigma(uv) \in K^*$. We obtain $uv \in L^*$ as desired. This completes the case where there are only finitely many c . In particular, we can derive $\text{SD}(A^*) = \text{AP}(A^*)$ at this point.

5.2 Infinitely many c

We have $w = cw'$ with $w' \in (B^*c)^\omega$. The substitution σ is extended from finite to infinite sequences $\sigma : (B^*c)^\omega \rightarrow T^\omega$ by $\sigma(v_1cv_2c\cdots) = [v_1]_\varphi[v_2]_\varphi\cdots$ for $v_i \in B^*$. By induction on the size of the monoid, there are languages $L_T, K_T \in \text{SD}(T^*)$ such that K_T is a prefix code of bounded synchronization delay, which satisfy:

$$\sigma(w') \in L_T K_T^\omega \subseteq \llbracket \sigma(w') \rrbracket_\psi.$$

The languages L_T and K_T can be chosen from a finite set depending on the homomorphism $\psi : T^* \rightarrow M_c$, only. Let $L = \sigma^{-1}(L_T)$ and $K = \sigma^{-1}(K_T)$. By Section 5.1 we know that $L, K \in \text{SD}(A^*)$ such that K is a prefix code of bounded synchronization delay. Let $L(w) = cLK^\omega$; then we have $w \in L(w)$. The product $L \cdot K^\omega$ is unambiguous provided that $L_T \cdot K_T^\omega$ is unambiguous. It remains to prove $cLK^\omega \subseteq \llbracket cw \rrbracket_\varphi$. This is achieved by some trick as done in [2].

Let $v \in cLK^\omega$. Then we have $\sigma(v') \in \llbracket \sigma(w') \rrbracket_\psi$ for $v = cv'$ and $w = cw'$. Hence $\sigma(v') \approx_\psi \sigma(w')$. Since \approx_ψ is the equivalence relation generated by \sim_ψ , it remains to show the following claim.

Claim: $\sigma(v) \sim_\psi \sigma(w)$ implies $cv \approx_\varphi cw$ for all $v, w \in (B^*c)^\omega$.

To see the claim write $\sigma(v) = \sigma(v_1c)\sigma(v_2c)\cdots$ and $\sigma(w) = \sigma(w_1c)\sigma(w_2c)\cdots$ with $v_i c, w_i c \in (B^*c)^+$ such that $\psi(\sigma(v_i c)) = \psi(\sigma(w_i c))$ for all $i \in \mathbb{N}$. It follows $\varphi(cv_i c) = \varphi(cw_i c)$ for all $i \in \mathbb{N}$, see Section 5.1. Thus

$$\begin{aligned} cv &= (cv_1c)v_2(cv_3c)v_4(c\cdots \\ &\sim_\varphi (cw_1c)v_2(cw_3c)v_4(c\cdots \\ &= cw_1(cv_2c)w_3(cv_4c)\cdots \\ &\sim_\varphi cw_1(cw_2)cw_3(cw_4c)\cdots \\ &= cw. \end{aligned}$$

This completes the proof of the claim and the proof of the theorem. \square

Corollary 1 *We have $\text{SD}(A^*) = \text{AP}(A^*)$ and $\text{SD}(A^\omega) = \text{AP}(A^\omega)$. Moreover, every language $\text{SD}(A^\omega)$ is a finite union of languages of the form LK^ω where $L, K \in \text{SD}(A^*)$ and K is a prefix code with bounded synchronization delay.*

Proof The first two statements immediately follow from $\text{SD}(A^\infty) = \text{AP}(A^\infty)$ in Theorem 1. The last property follows by taking $L_0 = \emptyset$ in statement 3 of the same theorem. \square

6 SD = AP implies SF = AP

The aim of this section is to show that Theorem 1 can be viewed as more general than the classic result that star-freeness is equivalent to aperiodicity. In this sense we would like to propose the thesis that $\text{SD} = \text{AP}$ is a more fundamental result in formal language theory than the celebrated result $\text{SF} =$

AP. We are aware that such a thesis is debatable. Therefore, we spend a few words to explain the idea.

First, by Theorem 1 we have $\text{SD}(A^\infty) = \text{AP}(A^\infty)$. Hence, $\text{SD}(A^\infty)$ is closed under complementation and therefore, trivially, $\text{SF}(A^\infty) \subseteq \text{SD}(A^\infty)$. Thus, in order to establish $\text{SD}(A^\infty) = \text{SF}(A^\infty) = \text{AP}(A^\infty)$ it remains to see that $\text{SD}(A^\infty) \subseteq \text{SF}(A^\infty)$. For this inclusion in turn, it is enough to prove the following simple fact.

Lemma 5 *If $K \in \text{SF}(A^*)$ is a prefix code of bounded synchronization delay, then $A^\infty \setminus K^\infty$ is star-free.*

Proof As K is a prefix code we can write $A^\infty \setminus K^\infty$ as an infinite union:

$$A^\infty \setminus K^\infty = \bigcup_{0 \leq i} (K^i A A^\infty \setminus K^{i+1} A^\infty). \quad (1)$$

Now, let d be the synchronization delay of K . Then we can write

$$A^\infty \setminus K^\infty = A^* K^d (A A^\infty \setminus K A^\infty) \cup \bigcup_{0 \leq i < d} (K^i A A^\infty \setminus K^{i+1} A^\infty).$$

The inclusion from left to right follows from Equation (1). The other inclusion holds since the intersection of K^∞ and $A^* K^d (A A^\infty \setminus K A^\infty)$ is empty. This is obtained by using the definition of synchronization delay. \square

Several comments are adequate.

Remark 1 The result of Lemma 5 for finitary languages is due to Schützenberger [15]. Another proof (which yields a star-free expression for $A^* \setminus K^*$) can be found in the textbook of Perrin and Pin [11, Lemma VIII.6.5]. \diamond

Remark 2 We have shown in Theorem 1 that requiring all products to be unambiguous is no restriction for the class SD; this still defines the same class of languages. When restricting star-free expressions over finite words in such a way that all products are unambiguous, then this yields a proper subclass of languages known as *unambiguous polynomials* [13, 16]. Its algebraic counterpart is the variety **DA**, see e.g. [17, 4]. In the proof of Lemma 5 this is reflected by the fact that, in general, the product $A^* \cdot K^d (A A^\infty \setminus K A^\infty)$ is not unambiguous. \diamond

Remark 3 The proof for $\text{SD}(A^\infty) = \text{SF}(A^\infty) = \text{AP}(A^\infty)$ can be arranged in a cycle. In this case we would start with Lemma 5 showing directly that $\text{SD}(A^\infty) \subseteq \text{SF}(A^\infty)$. Next, one uses the classical construction showing that $\text{AP}(A^\infty)$ is closed under concatenation. This yields $\text{SF}(A^\infty) \subseteq \text{AP}(A^\infty)$. Now, the final and most difficult step is just the implication from 2 to 3 in Theorem 1. This concludes the cycle. Note that in this cycle the statement of Lemma 4 has not been used because in a cycle it can be substituted by Lemma 5. \diamond

Remark 4 We have just argued that $SF(A^\infty) = AP(A^\infty)$ is an immediate consequence of the result $SD(A^\infty) = AP(A^\infty)$. We claim that it is however not equally simple to deduce $SD(A^\infty) = AP(A^\infty)$ from $SF(A^\infty) = AP(A^\infty)$. Indeed, the easy part is to see that $SD(A^\infty) \subseteq AP(A^\infty)$ or $SD(A^\infty) \subseteq SF(A^\infty)$. But then it remains the hard part which is to show one of the reverse inclusions. \diamond

Acknowledgements We would like to thank Jean-Éric Pin for bringing the class SD to our attention and for the proposal that the notion of *local divisor* might lead to a simplified proof for $SD(A^*) = AP(A^*)$.

References

1. A. Arnold. A syntactic congruence for rational ω -languages. *Theoretical Computer Science*, 39:333–335, 1985.
2. V. Diekert and P. Gastin. Pure future local temporal logics are expressively complete for Mazurkiewicz traces. *Information and Computation*, 204:1597–1619, 2006. Conference version in LATIN 2004, LNCS 2976, 170–182, 2004.
3. V. Diekert and P. Gastin. First-order definable languages. In *Logic and Automata: History and Perspectives*, Texts in Logic and Games, pages 261–306. Amsterdam University Press, 2008.
4. V. Diekert, P. Gastin, and M. Kufleitner. A survey on small fragments of first-order logic over finite words. *Int. J. Found. Comput. Sci.*, 19(3):513–548, 2008. Special issue DLT 2007.
5. V. Diekert, M. Kufleitner, and B. Steinberg. The Krohn-Rhodes theorem and local divisors. *Fundamenta Informaticae*, 116(1-4):65–77, 2012.
6. S. W. Golomb and B. Gordon. Codes with bounded synchronization delay. *Information and Control*, 8(4):355–372, 1965.
7. J. A. W. Kamp. *Tense Logic and the Theory of Linear Order*. PhD thesis, University of California, Los Angeles (California), 1968.
8. R. McNaughton and S. Papert. *Counter-Free Automata*. The MIT Press, 1971.
9. K. Meyberg. Lectures on algebras and triple systems. Technical report, University of Virginia, Charlottesville, 1972.
10. D. 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. D. Perrin and J.-É. Pin. *Infinite words*, volume 141 of *Pure and Applied Mathematics*. Elsevier, Amsterdam, 2004.
12. J.-É. Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.
13. J.-É. Pin, H. Straubing, and D. Thérien. Locally trivial categories and unambiguous concatenation. *Journal of Pure and Applied Algebra*, 52(3):297–311, 1988.
14. M. P. Schützenberger. On finite monoids having only trivial subgroups. *Inf. Control*, 8:190–194, 1965.
15. M. P. 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, London, 1975.
16. M. P. Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.
17. P. Tesson and D. Thérien. Diamonds are forever: The variety DA. In *Semigroups, Algorithms, Automata and Languages 2001, Proceedings*, pages 475–500. World Scientific, 2002.