Boolean Hierarchies inside Dot–Depth One

Heinz Schmitz*
Theoretische Informatik
Universität Würzburg
email: *schmitz@informatik.uni-wuerzburg.de*

Abstract

Let $\mathcal{B}_{1/2}$ denote the class of languages having dot–depth $1/2$, i.e., the class of languages that can we written as finite unions of languages $u_0A^+u_1A^+\cdots u_{n-1}A^+u_n$, where $u_i \in A^*$ and $n \geq 0$. A language has dot–depth one if and only if it is in the Boolean closure of $\mathcal{B}_{1/2}$. We examine the structure of the class of dot–depth one languages with respect to Boolean operations and identify an infinite family of Boolean hierarchies inside this class. In particular, we show that

1. the union of these hierarchies amounts to the Boolean hierarchy over $\mathcal{B}_{1/2}$,
2. all emerging inclusions are strict,
3. the membership problems for all classes in each hierarchy are decidable and
4. a given language can exactly be located in this landscape.

1 Introduction

We contribute to the study of starfree regular languages and their finite automata. Particularly, we deal with level one of the dot–depth hierarchy. This hierarchy is defined via alternating applications of concatenation and Boolean operations and exhausts the class of starfree languages. It was first studied in [CB71] and it is know to be strict [BK78]. Moreover, it is closely related to other research areas, e.g., connections to finite model theory, theory of finite semigroups, topology, Boolean circuits, complexity theory and others have been exposed. For an overview or as a good starting point to this rich field of research see e.g., [Brz76, Pin96a, Pin96b].

Let A be some finite alphabet with $|A| \geq 2$ and let $\mathcal{B}_{1/2}$ denote the class of languages having dot–depth $1/2$ (we use notations from [PW97]), i.e., the class of languages of A^+ that can we written as finite unions of languages $u_0A^+u_1A^+\cdots u_{n-1}A^+u_n$, where $u_i \in A^*$ and $n \geq 0$. A language has dot–depth one if and only if it is in the Boolean closure of $\mathcal{B}_{1/2}$, denoted as $\mathcal{B}_1 = \mathrm{BC}(\mathcal{B}_{1/2})$. To get to the starting point of our investigations let us recall what is known about the structure of $\mathcal{B}_1$.

The basic result was given by I. Simon in [Sim72], who characterized $\mathcal{B}_1$ in terms of a certain congruence $\sim_{m,k}$ with $k, m \geq 0$. Roughly speaking, two words x and y are equivalent under this congruence if they have the same first k letters, the same last k letters and if the same m-tuples of words of length $k+1$ appear in x and y. It is known that L is in $\mathcal{B}_1$ if and only if L can be written as a union of equivalence classes of $\sim_{m,k}$ for some integers m, k. Here the parameter m measures in a sense the number of concatenations in the product $u_0A^+\cdots u_{n-1}A^+u_n$ needed to define a language, where k corresponds to the maximal length of some u_i. For fixed integers m and k the obtained class is a Boolean algebra. An overview about the different subhierarchies resulting from fixing one or the other of these parameters is given in [Brz76].

A levelwise characterization of $\mathcal{B}_1$ was given by J. Stern in [Ste85a], who provided for $k \geq 0$ a description of classes $\mathcal{B}_{1,k}$ for which $\mathcal{B}_1 = \bigcup_{k\geq 0}\mathcal{B}_{1,k}$ by various means. First, it was shown that a

*Supported by the Deutsche Forschungsgemeinschaft (DFG), grant Wa 847/4-1.

language $L \subseteq A^+$ is in $\mathcal{B}_{1,k}$ if and only if there does not exist an infinite chain $w_1, w_2, w_3, \ldots$ of words where $w_i \preceq_k w_{i+1}$ and $w_i \in L \Leftrightarrow w_{i+1} \notin L$ for $i = 1, 2, 3, \ldots$. Here $\preceq_k$ denotes a certain generalization of the subword relation, where the k–prefix and the k–suffix are maintained, and where any letter from w_i appears in w_{i+1} and is followed by the same k letters in both words. Second, if $m_k^+(L)$ denotes the length of a maximal chain of the above kind starting with $w_1 \in L$, then also a bound on $m_k^+(L)$ in the alphabet size, automaton size and k was provided. Third, a forbidden pattern characterization of $\mathcal{B}_{1,k}$ was given, i.e., a deterministic finite automaton accepts a language from $\mathcal{B}_{1,k}$ if and only if certain patterns do not appear in the transition graph.

Moreover, since [PW97] a forbidden pattern characterization of $\mathcal{B}_{1/2}$ is known. Unfortunately, all these comprehensive results on the structure of $\mathcal{B}_1 = \mathrm{BC}(\mathcal{B}_{1/2})$ do not deal with the role of Boolean operations, an integral part in the definition of concatenation hierarchies. It is easy to see that $\mathcal{B}_{1,k}$ and $\mathcal{B}_{1/2}$ are incomparable for any k.

We take a closer look and reveal the fine structure of $\mathcal{B}_1$ with respect to the number of Boolean operations needed to describe a language — in comparison to the number k of consecutive letters that need to be specified. As a first step, we identify classes $\mathcal{B}_{1/2,k}$ such that $\mathcal{B}_{1/2} = \bigcup_{k \geq 0} \mathcal{B}_{1/2,k}$ and $\mathcal{B}_{1,k} = \mathrm{BC}(\mathcal{B}_{1/2,k})$. The definition of these classes is based on the fact that A^+ together with the relation $\preceq_k$ is a well partial ordered set. We give for all $k \geq 0$ a forbidden pattern characterization of $\mathcal{B}_{1/2,k}$, provide a formal language representation and show that each $\mathcal{B}_{1/2,k}$ is closed under union and intersection. A reproof of the mentioned forbidden pattern characterization for $\mathcal{B}_{1/2}$ follows.

For fixed $k \geq 0$ we look at the Boolean hierarchy over $\mathcal{B}_{1/2,k}$. Based on known normalform results for Boolean hierarchies we can relate particular values of m_k^+ to the single classes of the hierarchy. The hereby provided membership criterion can be applied in two ways. We derive a strictness argument and we show that for $l \geq 1$ and a given language L the question '$m_k^+(L) < l$' is decidable in nondeterministic logarithmic space. Hence, m_k^+ can be effectively computed with help of the bound from [Ste85a].

Then we give inclusion and strictness results for the classes of Boolean hierarchies for different values of k. In fact, for a given language L the graph of the monotone decreasing mapping $k \mapsto m_k^+(L)$ shows exactly the trade–off between k (hence, the number of consecutive letters) and the number of Boolean combinations one likes to spend when defining L in $\mathcal{B}_1$.

Finally, we relate our measures m_k^+ to the Boolean hierarchy over $\mathcal{B}_{1/2}$. This hierarchy was defined in [BKS98] and was very recently studied in [Gla99], where the decidability of the membership problems for its classes and its strictness were shown. A summary of structural results is given in Figure 7.

In case $k = 0$ we encounter (up to the empty word) the Boolean hierarchy over level $1/2$ of the so–called Straubing-Thérien hierarchy [Str81, Thé81, Str85], which was studied in [SW98].

2 Preliminaries

We make all definitions for an arbitrary finite alphabet A with $|A| \geq 2$. The set of all words over A is denoted by A^*, the set of all non-empty words over A is denoted by A^+, and the set of all words from A^* of length k is denoted by A^k for $k \geq 0$. Let ϵ denote the empty word and let $|x|$ be the length of a word $x \in A^*$. Throughout the paper (if no other definition is made) we consider languages as subsets of A^+. For $k \geq 0$ we denote by $A^{\leq k}$ ($A^{<k}$, $A^{\geq k}$, $\ldots$) the set of words from A^+ of length less or equal to k (less than k, greater or equal to k, $\ldots$ respectively). Note that none of these sets contains the empty word.

Let $x, y \in A^*$. We call y a *prefix* (*suffix*) of a word x if there exists a word z such that $x = yz$ ($x = zy$, resp.). For $k \geq 0$ the *k-prefix* and the *k-suffix* of $x = a_1 \cdots a_n$ are given by

$$p_k(x) =_{\mathrm{def}} \begin{cases} a_1 a_2 \cdots a_k, & \text{if } n \geq k, \\ x, & \text{otherwise} \end{cases}$$

and

$$s_k(x) =_{\text{def}} \begin{cases} a_{n-k+1}a_{n-k+2}\cdots a_n, & \text{if } n \geq k, \\ x, & \text{otherwise.} \end{cases}$$

We denote by x^R the reverse of a word x, i.e., $x^R =_{\text{def}} a_n a_{n-1} \cdots a_1$. For a language L let $L^R =_{\text{def}} \{ x^R \mid x \in L \}$. A deterministic finite automaton (dfa) F is given by $F = (A, S, \delta, s_0, S')$, where A is its input alphabet, S is its finite set of states, $\delta : A \times S \to S$ is its total transition function, $s_0 \in S$ is the starting state and $S' \subseteq S$ is the set of accepting states. For a dfa F we denote by $L(F)$ the language accepted by F. As usual we extend transition functions to input words, and we denote by $|F|$ the number of states of F. In this paper we will look only at dfa's where the starting state is not accepting, i.e., for any given dfa F we assume $L(F) \subseteq A^+$. We say that a state $s \in S$ has a loop $v \in A^+$ (has a v-loop, for short) if and only if $\delta(s, v) = s$.

For a class $\mathcal{C}$ of languages of A^+, let $\mathrm{BC}(\mathcal{C})$ be the Boolean closure of $\mathcal{C}$, i.e., $\mathrm{BC}(\mathcal{C})$ is the smallest class containing $\mathcal{C}$ and being closed under union, intersection and complementation, where the latter is taken with respect to A^+. For a class $\mathcal{C}$ of languages of A^+, let $\mathrm{POL}(\mathcal{C})$ be its polynomial closure defined as follows: A language $L \subseteq A^+$ is in $\mathrm{POL}(\mathcal{C})$ if and only if L can be written as a finite union of languages $u_0 L_1 u_1 L_2 u_2 \cdots L_n u_n$, where $u_i \in A^*$, $L_i \in \mathcal{C}$ and $n \geq 0$ (if $n = 0$ then it follows that $u_0 \neq \epsilon$). The dot-depth hierarchy can be defined as the following family of classes of languages of A^+:

1. $\mathcal{B}_0 =_{\text{def}} \{\emptyset, A^+\}$,

2. $\mathcal{B}_{n+1/2} =_{\text{def}} \mathrm{POL}(\mathcal{B}_n)$ for $n \geq 0$, and

3. $\mathcal{B}_{n+1} =_{\text{def}} \mathrm{BC}(\mathcal{B}_{n+1/2})$ for $n \geq 0$.

For a language $L \subseteq A^+$ and a minimal n with $L \in \mathcal{B}_{n/2}$ we say that L *has dot-depth* $n/2$. By definition, all $\mathcal{B}_{n+1/2}$ are closed under union and it is known, that these classes are also closed under intersection [Arf87, Arf91].

For a class $\mathcal{C}$ which is closed under union and intersection, the Boolean hierarchy over $\mathcal{C}$ is the family of classes $\mathcal{C}(k)$ and $\mathrm{co}\mathcal{C}(k)$ with $k \geq 1$, where $\mathcal{C}(k)$ can be defined (besides many other equivalent possibilities, cf. [KSW87, CGH+88]) as

$$\mathcal{C}(k) =_{\text{def}} \underbrace{\mathcal{C} \oplus \mathcal{C} \oplus \cdots \oplus \mathcal{C}}_{k \text{ times}},$$

where $\mathcal{C} \oplus \mathcal{C} =_{\text{def}} \{A \triangle B \mid A \in \mathcal{C},\ B \in \mathcal{C}\}$ and $\triangle$ denotes the symmetric set difference.

The following lemma states some well-known properties of the classes of the Boolean hierarchy over $\mathcal{C}$. Their normal form characterization in Statements 3 and 4 provides one of the other possibilities of their definition.

Lemma 2.1. *Let $\mathcal{C}$ be a class of languages which is closed under union and intersection, and let $l \geq 1$.*

1. $\mathrm{BC}(\mathcal{C}) = \bigcup_{l \geq 1} \mathcal{C}(l)$.

2. $\mathcal{C}(l) \cup \mathrm{co}\mathcal{C}(l) \subseteq \mathcal{C}(l+1) \cap \mathrm{co}\mathcal{C}(l+1)$.

3. $L \in \mathcal{C}(2l-1)$ *if and only if there exist languages* $L_1, L_2, \ldots, L_{2l-1} \in \mathcal{C}$ *such that* $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_{2l-1}$ *and* $L = \bigcup_{i=1}^{l-1}(L_{2i-1} \backslash L_{2i}) \cup L_{2l-1}$.

4. $L \in \mathcal{C}(2l)$ *if and only if there exist languages* $L_1, L_2, \ldots, L_{2l} \in \mathcal{C}$ *such that* $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_{2l}$ *and* $L = \bigcup_{i=1}^{l}(L_{2i-1} \backslash L_{2i})$.

It will be useful for us to look at a word $w \in A^+$ as a word over A^{k+1} by taking together each $k+1$ consecutive letters. We start with defining what the k-decomposition of a word is. In order to avoid confusion we will denote elements from A^{k+1} as $\alpha, \beta, \gamma, \ldots$ and subsets of A^{k+1} as $\Sigma, \Gamma, \ldots$. Let $k \geq 0$ and $w = a_1 a_2 \cdots a_{k+l} \in A^+$ for some $l \geq 1$. We call

$$\widehat{w} =_{\mathrm{def}} (\alpha_1, \alpha_2, \ldots, \alpha_l)$$

the *k-decomposition* of w if $\alpha_i = a_i \cdots a_{i+k}$ for $1 \leq i \leq l$. The value of k will always be clear from the context when we use the notation $\widehat{w}$. Intuitively, k indicates by how many letters from A consecutive α_i overlap. We set $\alpha(\widehat{w}) =_{\mathrm{def}} \{\alpha_1, \alpha_2, \ldots, \alpha_l\}$ as the set of elements from A^{k+1} occuring in the k-decomposition of w.

Next we want to define languages of words of A^+ that admit the same k-decomposition with respect to given elements and subsets of A^{k+1}. The definitions given here are a little more general than actually needed.

Definition 2.2. Let $k \geq 0$. Let $\alpha_1, \ldots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \Sigma_1, \ldots, \Sigma_n \subseteq A^{k+1}$ for some $n \geq 0$. For every $w \in A^+$ we say $w \in (\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n)_k$ if and only if $|w| \geq k+1$, $\widehat{w} = (\beta_1, \ldots, \beta_l)$ and there exist $0 = j_0 < j_1 < j_2 < \ldots < j_n < j_{n+1} = l+1$ such that

(a) $\beta_{j_i} = \alpha_i$, for $1 \leq i \leq n$ and
(b) $\beta_j \in \Sigma_i$ for $0 \leq i \leq n$ and $j_i < j < j_{i+1}$.

If we write the expression $(\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n)_k$ we understand this as a syntactical object describing some language. While being aware of this, we do not distinguish between such an object and the language it stands for, unless stated otherwise.

So the language $(\Sigma_0, \alpha_1, \Sigma_1, \alpha_2, \Sigma_2, \ldots, \alpha_n, \Sigma_n)_k$ consists of those words $w \in A^{\geq k+1}$, whose k-decomposition starts with a number (possibly zero) of elements from Σ_0, then α_1, followed by a number (possibly zero) of elements from Σ_1, then α_2 and so on. Note that such languages only contain words of a length that admits a k-decomposition and that in case $k = 0$ we deal with the usual concatenation, e.g. $(A_0, a_1, A_1, a_2, A_2)_0 = A_0^* a_1 A_1^* a_2 A_2^*$ and $(A_0)_0 = A_0^+$. We introduce another convenient notation.

Definition 2.3. Let $w, v \in A^*$, $\alpha_1, \ldots, \alpha_n \in A^{k+1}$ and $\Sigma_0, \ldots, \Sigma_n \subseteq A^{k+1}$ for some $n \geq 0$. Then we write $(w|\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n|v)_k$ instead of $\big(wA^* \cap A^*v \cap (\Sigma_0, \alpha_1, \Sigma_1, \ldots, \alpha_n, \Sigma_n)_k\big)$.

3 Known Characterizations of Languages of Dot-Depth One

We state characterizations of languages of dot-depth one to get to the starting point of our investigation. First we recall results on the structure of $\mathcal{B}_1$ that have been considered in the literature, and give them in terms of our notations and definitions.

The following definition of classes $\mathcal{B}_{1,(m,k)}$ is motivated by a characterization of $\mathcal{B}_1$ in terms of a certain congruence $\sim_{m,k}$ with $k, m \geq 0$ introduced by I. Simon in [Sim72]. Roughly speaking, two words x and y are equivalent under this congruence if they have the same first k letters, the same last k letters and if the same m-tuples of words of length $k+1$ appear in x and y.

Definition 3.1. Let $k \geq 0$ and $m \geq 1$. Then $\mathcal{B}_{1,(m,k)}$ is the class of languages $L \subseteq A^+$ that are in the Boolean algebra generated by languages L_i such that $L_i = D$ with $D \subseteq A^{<k+m}$ or

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $\alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$.

Note that we require here $m \geq 1$ which is no restriction since $(A^{k+1})_k$ contains only words of length $\geq k+1$ and hence $(A^{k+1})_k = \bigcup_{\beta \in A^{k+1}} (A^{k+1}, \beta, A^{k+1})_k$. It is known that increasing parameters k or m leads to refining congruences of $\sim_{m,k}$. Let, e.g., $L =_{\text{def}} (w|A^{k+1}, \alpha_1, A^{k+1}, \dots, \alpha_m, A^{k+1}|v)_k$. We observe that words $w \in L$ of length $k+m$ form a subset of $A^{<k+m+1}$ and that words $w \in L$ of length $> k+m$ have at least one position p in their k-decomposition which is not associated with an α_i. If we expand all α_i left (right) of p by one letter on the right side (left side, resp.), we obtain a representation with an increased k. If we explicitely add the element at position p to the description of L, we obtain a representation with an increased m. We conclude that $\mathcal{B}_{1,(m,k)}$ is included in $\mathcal{B}_{1,(m+1,k)} \cap \mathcal{B}_{1,(m,k+1)}$.

Lemma 3.2. *Let $k \geq 0$ and $m \geq 1$. Then it holds that*

1. $\mathcal{B}_{1,(m,k)} \subseteq \mathcal{B}_{1,(m,k+1)}$ *and*
2. $\mathcal{B}_{1,(m,k)} \subseteq \mathcal{B}_{1,(m+1,k)}$.

In fact, the proof can be carried out such that not full Boolean operations, but only finite unions are needed to express the increase in the parameters k and m. The next theorem establishes an important connection between $\mathcal{B}_1$ and the classes $\mathcal{B}_{1,(m,k)}$, i.e., the fact that the classes $\mathcal{B}_{1,(m,k)}$ refine $\mathcal{B}_1$. Originally, this theorem was stated for languages of A^* and in terms of the above mentioned congruence $\sim_{m,k}$ in [Sim72].

Theorem 3.3 [Sim72]. *Let $L \subseteq A^+$. Then $L \in \mathcal{B}_1$ if and only if there exist $k \geq 0$ and $m \geq 1$ such that $L \in \mathcal{B}_{1,(m,k)}$.*

For an overview on hierarchies which result from fixing one or the other parameter see [Brz76]. Among others, a hierarchy in $\mathcal{B}_1$ obtained by fixing the 'level' k of a language within this class has been studied.

Definition 3.4. Let $k \geq 0$. Then $\mathcal{B}_{1,k} =_{\text{def}} \bigcup_{m \geq 1} \mathcal{B}_{1,(m,k)}$.

The hierarchy of classes $\left\{ \mathcal{B}_{1,k} \;\middle|\; k \geq 0 \right\}$ is the so-called γ-hierarchy from [Brz76]. The following proposition is easy to see from Lemma 3.2.

Proposition 3.5. *It holds that*

1. $\mathcal{B}_{1,k} \subseteq \mathcal{B}_{1,k+1}$ *for each $k \geq 0$, and*
2. $\mathcal{B}_1 = \bigcup_{k \geq 0} \mathcal{B}_{1,k}$.

Various characterizations of the classes $\mathcal{B}_{1,k}$ were given by J. Stern in [Ste85a]. In order to restate them we need some more definitions, namely the notion of certain relations on words and particular subgraphs of the transition graph of a deterministic finite automaton. Let us start with a generalization of the well-known subword relation, which we extend to k-decompositions of words.

Definition 3.6. Let $k \geq 0$. Let $u, v \in A^{\geq k+1}$ and let $\widehat{u} = (\alpha_1, \alpha_2, \dots, \alpha_m)$ and $\widehat{v} = (\beta_1, \beta_2, \dots, \beta_n)$. We define

$$u \triangleleft_k v \iff_{\text{def}} \text{there exist } 1 \leq j_1 < j_2 \dots < j_m \leq n \text{ such that } \beta_{j_i} = \alpha_i.$$

Moreover, if $|u| \leq k$ we write $u \triangleleft_k v$ if and only if $u = v$.

If we additionally require that $p_k(u) = p_k(v)$ and $s_k(u) = s_k(v)$ we denote this by $u \preceq_k v$ and say that v is a k-extension from u. Note that both relations are the usual subword relation in case $k = 0$. Obviously, we have that $u \preceq_k v$ implies $u \triangleleft_k v$. A sequence of words $w_1, w_2, \ldots$ is called a $\triangleleft_k$–chain, if $w_i \triangleleft_k w_{i+1}$ for all words in the sequence (similarly, we define $\preceq_k$–chains). We say that a sequence has an alternation with respect to some given language L if $w_i \in L \iff w_{i+1} \notin L$ for some i.

Let $F = (A, S, \delta, s_0, S')$ be a dfa and let $k \geq 0$. Then F has pattern $\mathbf{T}_{\boldsymbol{k}}$ if there are states $s_1, \ldots, s_4, s, s' \in S$ and words $x, u, v, z \in A^*$ and $w \in A^k$ with $s \in S' \Leftrightarrow s' \notin S'$, $\delta(s_0, x) = s_1 = \delta(s_4, w)$, $\delta(s_1, u) = s_2$, $\delta(s_2, w) = s_3$, $\delta(s_3, v) = s_4$, $\delta(s_1, z) = s$ and $\delta(s_3, z) = s'$ (cf. Figure 1).

A dfa F has pattern $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ if there are states $s_1, \ldots, s_6, s, s' \in S$ and words $x, u, v, z \in A^*$ and $w \in A^k$ with $s \in S' \Leftrightarrow s' \notin S'$, $\delta(s_0, x) = s_1$, $\delta(s_1, w) = s_2$, $\delta(s_2, u) = s_3 = \delta(s_5, wu)$, $\delta(s_2, v) = s_4 = \delta(s_6, wv)$, $\delta(s_3, wv) = s_5$, $\delta(s_4, wu) = s_6$, $\delta(s_3, z) = s$ and $\delta(s_6, z) = s'$ (cf. Figure 2).

A dfa F has pattern $\mathbf{R}_{\boldsymbol{k}}$ if there are states $s_1, \ldots, s_8, s, s' \in S$ and words $x, y, y', u, v, z \in A^*$ and $w, w' \in A^k$ with $s \in S' \Leftrightarrow s' \notin S'$, $\delta(s_0, x) = s_1 = \delta(s_4, w)$, $\delta(s_1, y) = s_2$, $\delta(s_2, w') = s_3$, $\delta(s_3, y') = s_4$, $\delta(s_1, u) = s_5 = \delta(s_7, wu)$, $\delta(s_3, v) = s_6 = \delta(s_8, w'v)$, $\delta(s_5, w'v) = s_7$, $\delta(s_6, wu) = s_8$, $\delta(s_5, z) = s$ and $\delta(s_8, z) = s'$ (cf. Figure 3).

Theorem 3.7 [Ste85a]. *Let $L \subseteq A^+$, $k \geq 0$ and let F be the minimal dfa accepting L. Then the following statements are equivalent:*

(1) $L \in \mathcal{B}_{1,k}$

(2) F does not have any of the patterns $\mathbf{T}_{\boldsymbol{k}}$, $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ and $\mathbf{R}_{\boldsymbol{k}}$.

(3) any $\preceq_k$–chain has a finite number of alternations with respect to L

(4) in any $\preceq_k$–chain the number of alternations with respect to L is bounded by $2^{2|A|^{k+2}(k+1)^2|F|}$.

Let us comment the way we have cited here the original Theorem 3.3 from [Ste85a]. The relation 'k–embedding' used there is clearly intended to be the relation $\preceq_k$ for the following two reasons (although the definitions of both relations differ). First, the definition given in [Ste85a] contradicts Theorem 3.3 there, since it is required that u and v have the same $2k$-prefix if there is a k–embedding from u to v (see the definition on page 19 there). It is easy to see that for $k \geq 1$ and $L_{k+1} =_{\mathrm{def}} a^{k+1}A^+$ we have that $L_{k+1} \notin \mathcal{B}_{1,k}$ but any chain of k–embeddings has only a finite number of alternations with respect to L_{k+1}. Secondly, if one looks carefully at the proofs in [Ste85a] one can observe that in fact $\preceq_k$ is used.

In Statement (2) of Stern's Theorem it was required that F is k-stable and has no forks of type I and type II. Is is easy to see that the non-existence of Pattern $\mathbf{T}_{\boldsymbol{k}}$ is equivalent to the notion of k-stability defined there and that a fork of type I (type II) is exactly pattern $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ (pattern $\mathbf{R}_{\boldsymbol{k}}$, resp.) with the only diffence that it is not required that s_3 and s_6 (s_5 and s_8, respectively) are non-equivalent states, but they must lie in different strongly connected components (SCC). Let us argue why this is equivalent. First assume that F does not have any of the patterns $\mathbf{T}_{\boldsymbol{k}}$, $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ and $\mathbf{R}_{\boldsymbol{k}}$. Then F is k-stable and if the distinct states s_3 and s_6 in $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ lie in the same SCC, then we find $\mathbf{T}_{\boldsymbol{k}}$. The same holds if the distinct states s_5 and s_8 in $\mathbf{R}_{\boldsymbol{k}}$ are in one SCC. On the other hand assume that F is k-stable and has no forks of type I and II. Then F does not have $\mathbf{T}_{\boldsymbol{k}}$. Since s_3 and s_6 are in distinct SCC's they must be different and hence, they are not equivalent. The same holds for $\mathbf{R}_{\boldsymbol{k}}$. We did this little transformation to have a uniform description of the structure of F, from which we can now follow that $\mathbf{T}_{\boldsymbol{k}}$ and $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ are special cases of $\mathbf{R}_{\boldsymbol{k}}$.

Proposition 3.8. *Let $L \subseteq A^+$, $k \geq 0$ and let F be the minimal dfa accepting L. Then $L \in \mathcal{B}_{1,k}$ if and only if F does not have pattern $\mathbf{R}_{\boldsymbol{k}}$.*

Proof. Since the only-if part is clear by Theorem 3.7, we only have to argue that if F does not have pattern $\mathbf{R}_{\boldsymbol{k}}$ than it also does not have patterns $\mathbf{T}_{\boldsymbol{k}}$ and $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$. To get $\mathbf{T}_{\boldsymbol{k}}$ set $u = v = \epsilon$ and $w = w'$ in $\mathbf{R}_{\boldsymbol{k}}$ and to get $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ let $s_1 = s_3$, $y = y' = \epsilon$ and $w = w'$. ❑

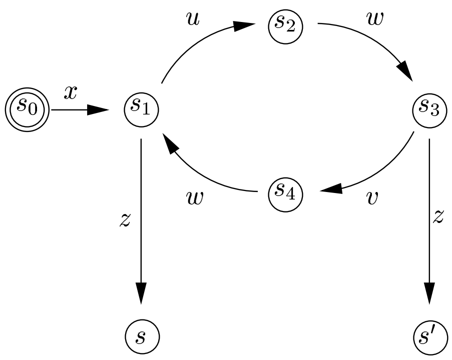

Figure 1: Pattern $\mathbf{T}_{\boldsymbol{k}}$ with $|w| = k$ and $s \in S' \Leftrightarrow s' \notin S'$.

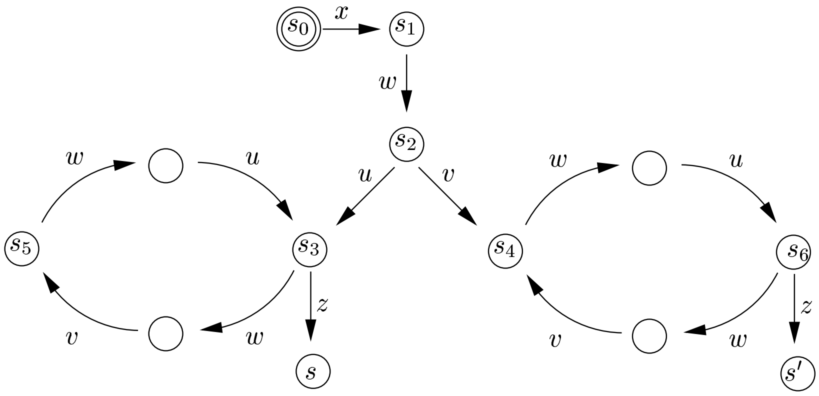

Figure 2: Pattern $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$ with $|w| = k$ and $s \in S' \Leftrightarrow s' \notin S'$.

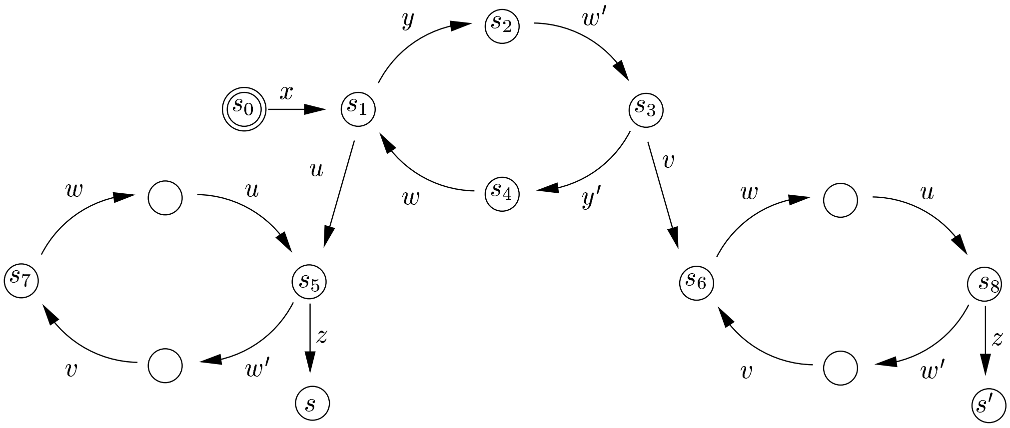

Figure 3: Pattern $\mathbf{R}_{\boldsymbol{k}}$ with $|w| = |w'| = k$ and $s \in S' \Leftrightarrow s' \notin S'$.

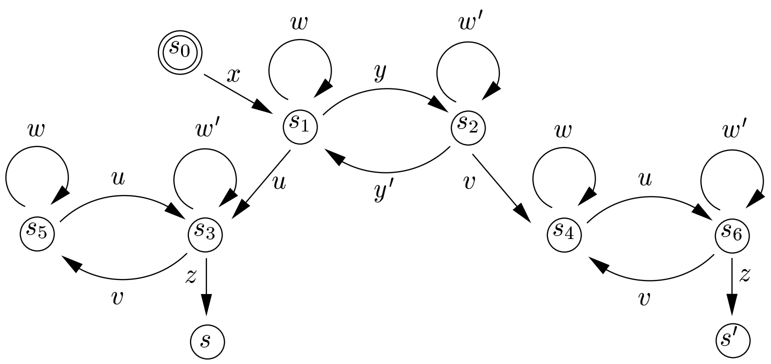

Figure 4: Forbidden pattern **R** for $\mathcal{B}_1$ with $|w|, |w'| \geq 1$ and $s \in S' \Leftrightarrow s' \notin S'$.

Finally, we want to mention that one can show with standard techniques that an arbitrary dfa F has one of the mentioned patterns if and only if the minimal dfa accepting $L(F)$ has the same pattern. Moreover, let $\hat{F}$ be a dfa with $L(\hat{F}) = L(F)^R$. Then F has pattern $\mathbf{T}_{\boldsymbol{k}}$ if and only if $\hat{F}$ has pattern $\mathbf{T}_{\boldsymbol{k}}^{\boldsymbol{rev}}$. It is known that the level k of a language L can be bounded.

Proposition 3.9 [Ste85a]. *Let F be a minimal dfa and set $N =_{\mathrm{def}} |F|^3$. Then $L(F) \in \mathcal{B}_1$ if and only if $L(F) \in \mathcal{B}_{1,N}$.*

The proof of this is straightforward and one only has to show the 'only-if'–part. Assume that $L(F) \notin \mathcal{B}_{1,N}$. Then there is pattern $\mathbf{R}_{\boldsymbol{N}}$ in F by Proposition 3.8. Now look at the $(N+1)$–triples of states induced by the prefixes of w starting with (s_4, s_6, s_7). Since there are only N different such triples, we encounter at least one triple twice. Hence, there is a factor $\hat{w} \in A^+$ of w which is in fact a $\hat{w}$–loop in all three occurences of w in $\mathbf{R}_{\boldsymbol{N}}$. In the same way we get a loop $\hat{w}' \in A^+$ as a factor of w'. Taking this together, we see that F has pattern $\mathbf{R}_{\boldsymbol{k}}$ for any $k \geq 0$ and hence $L(F) \notin \mathcal{B}_1$ by Proposition 3.5, a contradiction. In fact, this shows a forbidden pattern characterization of $\mathcal{B}_1$. Let us for this end make the following definition.

A dfa F has pattern **R** if there are states $s_1, \ldots, s_6, s, s' \in S$ and words $x, y, y', w, w', u, v, z \in A^*$ with $|w|, |w'| \geq 1$, $s \in S' \Leftrightarrow s' \notin S'$, $\delta(s_0, x) = s_1$, $\delta(s_1, y) = s_2$, $\delta(s_2, y') = s_1$, $\delta(s_i, w) = s_i$ for $i \in \{1, 4, 5\}$, $\delta(s_i, w') = s_i$ for $i \in \{2, 3, 6\}$, $\delta(s_1, u) = s_3 = \delta(s_5, u)$, $\delta(s_2, v) = s_4 = \delta(s_6, v)$, $\delta(s_3, v) = s_5$, $\delta(s_4, u) = s_6$, $\delta(s_2, z) = s$ and $\delta(s_5, z) = s'$ (cf. Figure 4).

Corollary 3.10. *Let $L \subseteq A^+$ and let F be the minimal dfa accepting L. Then $L \in \mathcal{B}_1$ if and only if F does not have pattern* **R**.

For a decision algorithm for $\mathcal{B}_1$ derived from the characterizations in [Ste85a] see [Ste85b, CH91]. The forbidden pattern **R** can also be derived from the algebraic characterization of the syntactic semigroups of dot–depth one languages given in [Kna83].

Finally, we recall in this section a forbidden pattern characterization of $\mathcal{B}_{1/2}$. Note that $\mathcal{B}_1 = \mathrm{BC}(\mathcal{B}_{1/2})$ and we will see that $\mathcal{B}_{1/2}$ and $\mathcal{B}_{1,k}$ are incomparable for any k (More precisely, $\mathcal{B}_{1,k} \not\subseteq \mathcal{B}_{1/2}$ can be followed from the proof of Theorem 7.7, and $\mathcal{B}_{1/2} \not\subseteq \mathcal{B}_{1,k}$ is clear from the proof of Proposition 8.2 below).

Let $F = (A, S, \delta, s_0, S')$ be a dfa. Then F has pattern **V** if there are states $s_1, s_2, s, s' \in S$ and words $x, w, z \in A^*$ and $v \in A^+$ with $\delta(s_0, x) = s_1 = \delta(s_1, w)$, $\delta(s_1, v) = s_2 = \delta(s_2, w)$, $\delta(s_1, z) = s \in S'$ and $\delta(s_2, z) = s' \notin S'$, (cf. Figure 5).

Theorem 3.11 [PW97]. *Let F be a minimal dfa with $L(F) \subseteq A^+$. Then $L(F) \in \mathcal{B}_{1/2}$ if and only if F does not have pattern* **V**.

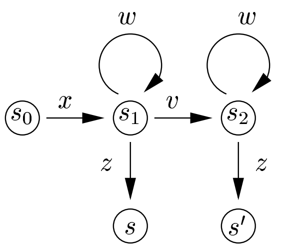

Figure 5: Forbidden pattern **V** for $\mathcal{B}_{1/2}$ with $w \in A^+$, $v, z \in A^*$ and $s \in S', s' \notin S'$.

4 Well Partial Orderings

As is turn out, the defined generalizations of the subword relation share an interesting property which gives rise to further investigations. Let $k \geq 0$. We show that A^+ together with $\triangleleft_k$ is a well partial ordered set. The same holds for $\preceq_k$ since restrictions of constant size in a word do not affect this notion.

For several equivalent properties, which may be used for defining well partial orderings, see [CK96, SS83]. We will show here that in A^+ there exists neither an infinite strictly descending $\triangleleft_k$–chain of words nor an infinite set of pairwise incomparable elements with respect to $\triangleleft_k$. By [CK96] this is equivalent to saying that for every nonempty subset of A^+ the set of minimal elements with respect to $\triangleleft_k$ in this subset is nonempty and finite.

First observe, by a typical length argument (namely that $u \triangleleft_k v$ with $u \neq v$ implies $|u| < |v|$), that we only have to show that any set of pairwise incomparable elements is finite. We will make use of the following arguments in the proof below.

Proposition 4.1. *Let $k \geq 0$, $w \in A^k$ and $a \in A$. Then it holds that*

1. *$u \triangleleft_k au$ for every $u \in A^{>k}$, and*
2. *$wx \triangleleft_k wy \implies awx \triangleleft_k awy$ for every $x, y \in A^*$.*

Proof. Statement 1 is obvious. To see 2 note that if $x = \epsilon$ then it must be that $y = \epsilon$. If $x \neq \epsilon$ then we only need to observe that aw as the first element of $\widehat{awx}$ appears also as the first element in $\widehat{awy}$. ❑

We prove an extension of a fundamental theorem from Higman [Hig52], which we encounter in case $k = 0$ (i.e., the subword relation, also called division ordering).

Theorem 4.2. *Let $k \geq 0$. Any subset of A^+ of pairwise incomparable words w.r.t. $\triangleleft_k$ is finite.*

Proof. We adapt the proof for the subword relation as stated in [SS83] for our purposes. So assume that there are infinite subsets of A^+ all consisting of pairwise incomparable words w.r.t. $\triangleleft_k$. In particular there exist infinite sequences $\{f_i\}$ of words such that from $i < j$ it follows that $f_i \not\triangleleft_k f_j$. We will show that this is not true. For this consider any such sequence $\{f_i\}$ and note that all words in such a sequence must be different since $\triangleleft_k$ is reflexive. We choose from these sequences an 'earliest' sequence $\{u_i\}$ as follows: Let u_1 be a shortest word beginning some sequence $\{f_i\}$, then let u_2 be a shortest second word of any sequence $u_1, f_2, f_3 \ldots$, then let u_3 be a shortest third word of any sequence $u_1, u_2, f_3 \ldots$, and so on. Clearly, also for $\{u_i\}$ it holds that from $i < j$ it follows $u_i \not\triangleleft_k u_j$. Since we have a finite alphabet there are words $u_{i_1} = awg_1, u_{i_2} = awg_2, \ldots$ with $i_1 < i_2 < \ldots$ for some $a \in A$ and some $w \in A^k$. Note that none of these g_j can be ϵ since $aw \triangleleft_k awv$ for arbitrary v.

Now we look at the sequence $u_1, u_2, \ldots, u_{i_1 - 1}, wg_1, wg_2, \ldots$ Denote this new sequence as $\{x_i\}$ which is 'earlier' than $\{u_i\}$ since $|wg_1| < |u_{i_1}|$. In order to obtain a contradiction to our construction we need to show that for all i, j with $i < j$ we can conclude $x_i \not\triangleleft_k x_j$. If $i, j \in \{1, \ldots, i_1 - 1\}$ then this is clear by the same property for $\{u_i\}$. Now suppose $i \in \{1, \ldots, i_1 - 1\}$ and $j \geq i_1$ and assume $x_i \triangleleft_k x_j$ where $x_i = u_i$ and $x_j = wg_l$ for some $l \geq 1$. Since $wg_l \in A^{>k}$ we have by Proposition 4.1.1 that $wg_l \triangleleft_k awg_l = u_{i_l}$ and together $u_i \triangleleft_k u_{i_l}$, a contradiction. Finally let $i, j \geq i_1$ with $i < j$. Assume $x_i \triangleleft_k x_j$ with $x_i = wg_l$ and $x_j = wg_m$ for some $l < m$. By Proposition 4.1.2 it follows that $awg_l \triangleleft_k awg_m$, so $u_{i_l} \triangleleft_k u_{i_m}$, again a contradiction. ❑

Theorem 4.3. *Let $k \geq 0$. Any subset of A^+ of pairwise incomparable words w.r.t. $\preceq_k$ is finite.*

Proof. Suppose there exist infinite subsets of A^+ of pairwise incomparable words w.r.t. $\preceq_k$. Then there is also an infinite subset L such that all words in L have the same k-prefix and k-suffix. So the words in L are pairwise incomparable w.r.t. $\triangleleft_k$ contradicting Theorem 4.2. ❑

5 The Structure inside $\mathcal{B}_1$

We take a detailed look inside $\mathcal{B}_1$. First, we define the set of k-extensions from a word u as $\langle u \rangle_k =_{\mathrm{def}} \{ v \mid u \preceq_k v \}$. We can describe this set over A^{k+1}.

Proposition 5.1. *Let $k \geq 0$. If $u \in A^{\leq k}$ then $\langle u \rangle_k = \{u\}$. If $u \in A^{\geq k+1}$ and $\widehat{u} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ then*

$$\langle u \rangle_k = \Big(p_k(u) | A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1} | s_k(u) \Big)_k$$

Proof. Clearly, any word from the right hand side is a k-extension of u. On the other hand, any k-extension of u must have u's k-prefix and k-suffix and all elements of u's k-decomposition must occur in this ordering. ❑

The following duality holds.

Proposition 5.2. *For $k \geq 0$ and $u, v \in A^+$ it holds that $u \preceq_k v \iff \langle u \rangle_k \supseteq \langle v \rangle_k$.*

The implication from left to right is due to the transitivity of $\preceq_k$. The other implication follows from its reflexivity. Next, we give another description of the languages in $\mathcal{B}_{1,k}$ over A^{k+1}. The important difference to Definition 3.4 is here that we can get rid of the parameter m occuring in the size of the finite set D. The main argument comes from the fact that $\preceq_k$ is a well partial ordering. We write $\langle L \rangle_k =_{\mathrm{def}} \bigcup_{u \in L} \langle u \rangle_k$ for the set of all words which are k-extensions of some word in a given language L.

Theorem 5.3. *Let $k \geq 0$ and $L \subseteq A^+$. Then L is in $\mathcal{B}_{1,k}$ if and only if L is in the Boolean algebra generated by languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k}$ or*

$$L_i = (w | A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1} | v)_k$$

where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$.

Proof. We see that the languages L_i mentioned in the theorem are in classes $\mathcal{B}_{1,(m,k)}$ for a finite number of different values of m. By Lemma 3.2 we may choose a maximal such m and perform the Boolean operations in the respective class. This shows the if-part.

To see the only if-part let $L \in \mathcal{B}_{1,k}$ which is by definition and Lemma 3.2 a Boolean combination of languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k+m_i}$ for some $m_i \geq 1$ or

$$L_i = (w_i | A^{k+1}, \alpha^i_1, A^{k+1}, \alpha^i_2, A^{k+1}, \ldots, \alpha^i_{m_i}, A^{k+1} | v_i)_k$$

with $m_i \geq 1, \alpha^i_1, \ldots, \alpha^i_{m_i} \in A^{k+1}$ and $w_i, v_i \in A^k$. We only have to care about the finite sets D_i, i.e., we have to show that we can express the words $u \in D_i$ with $|u| > k$ by Boolean combinations of languages of the form $(w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_l, A^{k+1}|v)_k$ for some $l \geq 1, \alpha_1, \ldots, \alpha_l \in A^{k+1}$ and $w, v \in A^k$.

Observe that $u \in \langle u \rangle_k$ which has the desired form by Proposition 5.1. We want to show that we can subtract all other words in $\langle u \rangle_k$, i.e., describe $S =_{\mathrm{def}} \langle u \rangle_k \backslash \{u\}$ by a Boolean combination of the requested type. Let $y \in S$ and $x \in \langle S \rangle_k$ with $y \preceq_k x$. Then $u \preceq_k y$ and $y \neq u$. It follows that $u \preceq_k y \preceq_k x$ and $x \neq u$ (by a length argument), hence $S = \langle S \rangle_k$. Now we can write

$$S = \langle S \rangle_k = \bigcup_{x \in S} \langle x \rangle_k$$

and Theorem 4.3 tells us that this union is in fact finite. ❑

So each class $\mathcal{B}_{1,k}$ can be understood as the Boolean closure of the class of all languages having the form from the previous theorem. We name the class of these languages as $\mathcal{B}_{1/2,k}$ in analogy to $\mathcal{B}_1 = \mathrm{BC}(\mathcal{B}_{1/2})$.

Definition 5.4. Let $k \geq 0$. A language $L \subseteq A^+$ belongs to $\mathcal{B}_{1/2,k}$ if and only if it can be written as a finite union of languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k}$ or

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$.

We think that the classes $\mathcal{B}_{1/2,k}$ fit in a natural way into the landscape of classes inside dot-depth one. As a further application of Theorem 4.3 we obtain an interesting property of these classes, i.e., we see that the closure of an arbitrary language under k-extensions is in $\mathcal{B}_{1/2,k}$. In particular, we obtain a regular language, independent from the languages we start with.

Theorem 5.5. *Let $k \geq 0$ and $L \subseteq A^+$. Then $\langle L \rangle_k \in \mathcal{B}_{1/2,k}$.*

Proof. By definition, $\langle L \rangle_k = \bigcup_{u \in L} \langle u \rangle_k$ and by Proposition 5.1 we have $\langle u \rangle_k \in \mathcal{B}_{1/2,k}$. By Theorem 4.3 we conclude that only a finite number of different elements from L strictly contribute to the union. ❑

With the next two propositions, we give some closure properties of the classes $\mathcal{B}_{1/2,k}$ and relate them to the classes known from literature.

Proposition 5.6. *Let $k \geq 0$.*

1. *Let $L \subseteq A^+$. Then $L \in \mathcal{B}_{1/2,k}$ if and only if $L = \langle L \rangle_k$.*
2. *The classes $\mathcal{B}_{1/2,k}$ and $\mathrm{co}\mathcal{B}_{1/2,k}$ are closed under finite union and intersection.*

Proof. To see Statement 1 first suppose $L = \langle L \rangle_k$. Then by Theorem 5.5 we have $L \in \mathcal{B}_{1/2,k}$. Conversely, if $u \in \langle L \rangle_k$ then there is some $x \in L$ such that $x \preceq_k u$. We need to show $u \in L$. If $|x| \leq k$ then $u = x$. If $|x| > k$ then x is in some set of the form $(w|A^{k+1}, \alpha_1, \ldots, \alpha_m, A^{k+1}|v)_k$ where $m \geq 0, \alpha_1, \ldots, \alpha_m \in A^{k+1}$ and $w, v \in A^k$. So the k-prefix of x is w, the k-suffix of x is v and $\alpha_1, \ldots, \alpha_m$ appear in the k-decomposition of x. By the definition of $\preceq_k$ then u also has k-prefix w and k-suffix v and all elements in $\widehat{x}$ appear in this order in $\widehat{u}$. In particluar, $\alpha_1, \ldots, \alpha_m$ appear in the k-decomposition of u, so $u \in L$.

We turn to Statement 2. The closure of $\mathcal{B}_{1/2,k}$ under union is by definition. So take two languages from $\mathcal{B}_{1/2,k}$ each of which being a finite union of languages L_i such that $L_i = D$ with $D \subseteq A^{\leq k}$ or

$$L_i = (w|A^{k+1}, \alpha_1, A^{k+1}, \alpha_2, A^{k+1}, \ldots, \alpha_m, A^{k+1}|v)_k$$

where $m \geq 1, \alpha_1, \ldots, \alpha_m \in A^{k+1}$, $w, v \in A^k$, and look at their intersection. First we may distribute the intersection over all unions. Clearly, the intersection of two finite sets D as above is again a subset of $A^{\leq k}$. If we have to consider mixed terms with a set D and an infinite set then the intersection is empty. So we only have to care about intersections $L_1 \cap L_2$ where $L_i = (w_i|A^{k+1}, \alpha_1^i, A^{k+1}, \ldots, \alpha_{m_i}^i, A^{k+1}|v_i)_k$ for $i = 1, 2$. If $w_1 \neq w_2$ or $v_1 \neq v_2$ then again the intersection is empty. So we may assume $w_1 = w_2$ and $v_1 = v_2$ and we state that $L_1 \cap L_2 = L'$ with

$$L' =_{\text{def}} \bigcup (w_1|A^{k+1}, \beta_1, A^{k+1}, \ldots, A^{k+1}, \beta_m, A^{k+1}|v_1)_k$$

where the union goes over all tuples $(\beta_1, \ldots, \beta_m)$ with $1 \leq m \leq m_1 + m_2$ and $\beta_i \in A^{k+1}$ for $1 \leq i \leq m$ such that we find $(\alpha_1^1, \ldots, \alpha_{m_1}^1)$ and $(\alpha_1^2, \ldots, \alpha_{m_2}^2)$ in this tuple. Hence there exist $1 \leq i_1 < i_2 < \ldots < i_{m_1} \leq m$ and $1 \leq j_1 < j_2 < \ldots < j_{m_2} \leq m$ such that $(\alpha_1^1, \alpha_2^1, \ldots, \alpha_{m_1}^1) = (\beta_{i_1}, \beta_{i_2}, \ldots, \beta_{i_{m_1}})$ and $(\alpha_1^2, \alpha_2^2, \ldots, \alpha_{m_2}^2) = (\beta_{j_1}, \beta_{j_2}, \ldots, \beta_{j_{m_2}})$.

Observe that $L' \in \mathcal{B}_{1/2,k}$ since m is bound by a constant. In the k-decomposition of each word $u \in L'$ we find by the choice of $(\beta_1, \ldots, \beta_m)$ the required elements of A^{k+1} to see that u is in L_1 and also in L_2, so $u \in L_1 \cap L_2$.

Now let $u \in L_1 \cap L_2$ and $\widehat{u} = (\gamma_1, \ldots, \gamma_l)$ for some $l \geq 1$. Thus there must be sequences of indices $1 \leq i_1 < i_2 < \ldots < i_{m_1} \leq l$ and $1 \leq j_1 < j_2 < \ldots < j_{m_2} \leq l$ such that $(\alpha_1^1, \ldots, \alpha_{m_1}^1) = (\gamma_{i_1}, \ldots, \gamma_{i_{m_1}})$ and $(\alpha_1^2, \ldots, \alpha_{m_2}^2) = (\gamma_{j_1}, \ldots, \gamma_{j_{m_2}})$. Then there is also a sequence of indices $1 \leq l_1 < l_2 < \ldots < l_m \leq l$ with $m \leq m_1 + m_2$ and $\{l_1, \ldots, l_m\} = \{i_1, \ldots, i_{m_1}, j_1, \ldots, j_{m_2}\}$. So $(\gamma_{l_1}, \gamma_{l_2}, \ldots, \gamma_{l_m})$ is one of the tuples $(\beta_1, \ldots, \beta_m)$ occuring in the definition of L' and we see that $u \in L'$. ❑

Proposition 5.7. *It holds that*

1. $\mathcal{B}_{1,k} = \mathrm{BC}(\mathcal{B}_{1/2,k})$ *for each* $k \geq 0$,
2. $\mathcal{B}_{1/2,k} \subseteq \mathcal{B}_{1/2,k+1}$ *for each* $k \geq 0$, *and*
3. $\mathcal{B}_{1/2} = \bigcup_{k \geq 0} \mathcal{B}_{1/2,k}$.

Statement 1 is due to Theorem 5.3 which gave rise to the definition of $\mathcal{B}_{1/2,k}$. Statement 2 can be followed from the way Lemma 3.2 can be proved (see the remark following there). In case of Statement 3 there is in fact something to do. Again, we omit a proof here since it can be followed from the way Theorem 3.3 can be proved.

6 Elementary k-Extensions and a Forbidden Pattern Characterization of $\mathcal{B}_{1/2,k}$

In this section we prove a forbidden pattern theorem for the classes $\mathcal{B}_{1/2,k}$ using the Pattern $\mathbf{V}_k$ (cf. Figure 6). Let F be a dfa accpeting some language L. We prove that a language L belongs to $\mathcal{B}_{1/2,k}$ if and only if we can not find Pattern $\mathbf{V}_k$ in the transition graph of F. Before that, we introduce for technical reasons the notion of an elementary k-extension and show that we can decompose any arbitrary k-extension into a finite number of elementary ones. This together with results from the previous section will lead to a short proof of the aspired forbidden pattern theorem.

We can view an arbitrary k-extension $u \preceq_k v$ as a simultaneous insertion of factors at different positions in u to obtain v, while respecting certain context conditions. If there is only one such insertion position, we call this an elementary k-extension from u to v.

Definition 6.1. Let $k \geq 0$. Let $u, v \in A^{\geq k+1}$ and let $\widehat{u} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $\widehat{v} = (\beta_1, \beta_2, \ldots, \beta_n)$ for some $m, n \geq 1$. We define

$$u \preceq_k^{\mathrm{e}} v \iff_{\mathrm{def}} \begin{array}{ll} \text{(a)} & p_k(u) = p_k(v), s_k(u) = s_k(v) \text{ and,} \\ \text{(b)} & \text{there exist } r \geq 1 \text{ with } n = m + r \text{ and } 0 \leq l \leq m \text{ such that} \\ & \quad (\beta_1, \ldots, \beta_n) = (\alpha_1, \ldots, \alpha_l, \gamma_1, \ldots, \gamma_r, \alpha_{l+1}, \ldots, \alpha_m) \\ & \text{for some } \gamma_1, \ldots, \gamma_r \in A^{k+1} \end{array}$$

As in the case of $\preceq_k$, if $|u| \leq k$ we write $u \preceq_k^{\mathrm{e}} v$ if and only if $u = v$.

We call this an elementary k-extension. Clearly, if $u \preceq_k^{\mathrm{e}} v$ then $u \preceq_k v$. Note also that if $u \preceq_k^{\mathrm{e}} v$ then there are $x, y, z \in A^*$ with $u = xz$ and $v = xyz$.

Proposition 6.2. *Let $k \geq 0$ and let $u, v \in A^+$. If $u \neq v$ and $u \preceq_k v$ then there exists some $w \in A^+$ with $w \neq u$ and $u \preceq_k^{\mathrm{e}} w \preceq_k v$.*

Proof. We may assume $u \preceq_k v$ with $u, v \in A^{>k}$. Let $\widehat{u} = (\alpha_1, \ldots, \alpha_m)$ and $\widehat{v} = (\beta_1, \ldots, \beta_n)$ for some $m, n \geq 1$. By definition $p_k(u) = p_k(v)$ and $s_k(u) = s_k(v)$ and there are $1 \leq j_1 < j_2 \ldots < j_m \leq n$ such that $\beta_{j_i} = \alpha_i$ for $1 \leq i \leq m$. Set $j_0 = 0$ and $j_{m+1} = n + 1$. Fix now a maximal l with $0 \leq l \leq m$ such that $j_l = l$ and set $r =_{\mathrm{def}} j_{l+1} - j_l - 1$. Note that such an l exists since $u \neq v$, and that $r \geq 1$. We claim that there is a word $w \in A^{>k}$ such that

$$\widehat{w} = (\alpha_1, \ldots, \alpha_l, \beta_{l+1}, \ldots, \beta_{l+r}, \alpha_{l+1}, \ldots, \alpha_m).$$

If we can show that such a word indeed exists, we immediately have $u \preceq_k^{\mathrm{e}} w$. Only a few observations are needed. To see that $s_k(\alpha_l) = p_k(\beta_{l+1})$ note that $l = j_l$ and thus $\alpha_l = \beta_l$. To see that $s_k(\beta_{l+r}) = p_k(\alpha_{l+1})$ observe that $l + r + 1 = j_{l+1}$ and thus $\alpha_{l+1} = \beta_{l+r+1}$. If $0 < l < m$ then $p_k(w) = p_k(u)$ and $s_k(w) = s_k(u)$. If $l = 0$ then $p_k(w) = p_k(v)$ and hence $p_k(w) = p_k(u)$. The same holds for the k-suffix if $l = m$.

We need to show $w \preceq_k v$. In fact we only need to show $w \triangleleft_k v$. Due to the construction of w we may take as a witnessing sequence of indices $1, 2, \ldots, l, l+1, \ldots, l+r, j_{l+1}, j_{l+2}, \ldots, j_m$. Recall that $j_{l+1} = l + r + 1$. ❑

Proposition 6.3. *Let $k \geq 0$ and let $u, v \in A^+$ with $u \neq v$ and $u \preceq_k v$. Then there exist $r \geq 1$ and words $w_0, w_1, \ldots, w_r \in A^+$ with $w_0 = u, w_r = v$ and for all $0 \leq i < r$ it holds that $w_i \preceq_k^{\mathrm{e}} w_{i+1}$.*

Proof. Define $w_0 =_{\mathrm{def}} u$. If v is already an elementary k-extension from u we are done with $w_1 =_{\mathrm{def}} v$. Otherwise we apply Proposition 6.2, set $w_1 =_{\mathrm{def}} w$ and obtain $w_0 \preceq_k^{\mathrm{e}} w_1$. Now we start over again with w_1 and v, apply Proposition 6.2 if necessary, and so on. This procedure comes to an end since Proposition 6.2 provides a strict elementary k-extension and strict extensions are length increasing. ❑

We define a family of patterns in transition graphs of dfa's which characterizes the classes $\mathcal{B}_{1/2,k}$. Let $F = (A, S, \delta, s_0, S')$ be a dfa and let $k \geq 0$. Then F has pattern $\mathbf{V}_{\boldsymbol{k}}$ if there are states $s_1, s_2, s, s' \in S$ and words $x, v, z \in A^*$ with $|x| \geq k$, $|xz| \geq k + 1$, $s_k(x) = s_k(xv)$ and $\delta(s_0, x) = s_1$, $\delta(s_1, v) = s_2$, $\delta(s_1, z) = s \in S'$ and $\delta(s_2, z) = s' \notin S'$ (cf. Figure 6).

Theorem 6.4. *Let $k \geq 0$ and $L \subseteq A^+$. Let $F = (A, S, \delta, s_0, S')$ be a dfa with $L(F) = L$. Then $L \in \mathcal{B}_{1/2,k}$ if and only if F does not have pattern $\mathbf{V}_{\boldsymbol{k}}$.*

Proof. Suppose a dfa F has pattern $\mathbf{V}_{\boldsymbol{k}}$ witnessed by x, v and z as in the patterns definition. We see that $xz \preceq_k xvz$, because both words have the same k-prefix and k-suffix, and each element of $\widehat{xz}$ appears in $\widehat{xvz}$ since $s_k(x)z = s_k(xv)z$. From $xz \in L$ and $xvz \notin L$ we conclude that L is not closed under k-extensions. With Proposition 5.6 we obtain $L \notin \mathcal{B}_{1/2,k}$.

$$\begin{tikzcd} s_0 \arrow[r, "x"] & s_1 \arrow[r, "v"] \arrow[d, "z"'] & s_2 \arrow[d, "z"] \\ & s & s' \end{tikzcd}$$

Figure 6: Forbidden pattern $\mathbf{V}_{\boldsymbol{k}}$ for $\mathcal{B}_{1/2,k}$ with $x, v, z \in A^*$,$|x| \geq k$, $|xz| \geq k+1$, $s_k(x) = s_k(xv)$ and $s \in S', s' \notin S'$.

On the other hand assume that $L \neq \langle L \rangle_k$, i.e., by Proposition 5.6 there are words $u \in L$ and $v \notin L$ with $u \preceq_k v$. By Proposition 6.3 there is some $r \geq 1$ and words $w_0, w_1, \ldots, w_r \in A^+$ with $w_0 = u, w_r = v$ and for all $0 \leq i < r$ it holds that $w_i \preceq_k^{\mathrm{e}} w_{i+1}$. Moreover, there must be some i with $w_i \in L$ and $w_{i+1} \notin L$ and $w_i, w_{i+1} \in A^{\geq k+1}$. Let $\widehat{w_i} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $\widehat{w}_{i+1} = (\alpha_1, \ldots, \alpha_l, \gamma_1, \ldots, \gamma_r, \alpha_{l+1}, \ldots, \alpha_m)$ for appropriate $0 \leq l \leq m$ and $\gamma_1, \ldots, \gamma_r \in A^{k+1}$ for some $r \geq 1$. Let $a_1, \ldots, a_l$ be the first letters of $\alpha_1, \ldots, \alpha_l$, respectively, and let $b_1, \ldots, b_r, c_{l+1}, \ldots, c_m$ be the last letters of $\gamma_1, \ldots, \gamma_r, \alpha_{l+1}, \ldots, \alpha_m$, respectively. Set $x =_{\mathrm{def}} a_1 \cdots a_l s_k(\alpha_l)$, $v =_{\mathrm{def}} b_1 \cdots b_r$ and $z =_{\mathrm{def}} c_{l+1} \cdots c_m$. Then $|x| \geq k$, $xz = w_i$, $|xz| \geq k+1$ and $xvz = w_{i+1}$. It holds that $s_k(x) = p_k(\alpha_{l+1}) = s_k(\gamma_r) = s_k(xv)$ if $0 < l < m$. If $l = 0$ then $x = p_k(\gamma_1) = p_k(\alpha_1) = s_k(xv)$ since $p_k(w_i) = p_k(w_{i+1})$. Analogously, if $l = m$ then $s_k(x) = s_k(\alpha_l) = s_k(\gamma_r) = s_k(xv)$ since $s_k(w_i) = s_k(w_{i+1})$. Together we see that x, v and z give rise to pattern $\mathbf{V}_{\boldsymbol{k}}$ in F. ❑

The previous theorem yields for fixed k an NL–algorithm to decide for a given dfa F the question '$L(F) \in \mathcal{B}_{1/2,k}$'. We only have to look for Pattern $\mathbf{V}_{\boldsymbol{k}}$ in the transition graph of F.

Proposition 6.5. *Let F be a dfa. Then F does not have pattern* $\mathbf{V}$ *if and only if there exists some $k \geq 0$ such that F does not have pattern* $\mathbf{V}_{\boldsymbol{k}}$*.*

Proof. Let $F = (A, S, \delta, s_0, S')$. To see the if-part suppose there is some $k \geq 0$ such that F does not have pattern $\mathbf{V}_{\boldsymbol{k}}$, but F does have pattern $\mathbf{V}$ witnessed by x, w, v and z. Then F also has pattern $\mathbf{V}_{\boldsymbol{l}}$ with $l =_{\mathrm{def}} |w|^k$ witnessed by $x' =_{\mathrm{def}} xw^k$, $v' =_{\mathrm{def}} vw^k$ and z. Since $l \geq k$ there is also pattern $\mathbf{V}_{\boldsymbol{k}}$ in F, a contradiction.

For the other implication assume that F does not have pattern $\mathbf{V}$, but for all $k \geq 0$ there is pattern $\mathbf{V}_{\boldsymbol{k}}$ in F, witnessed by x_k, v_k and z_k. Especially there are states $s_1, s_2, s_3, s_4, s, s' \in S$ involved in pattern $\mathbf{S}_{\boldsymbol{l}}$ for $l =_{\mathrm{def}} |S|^2 + 1$ in the following way: There are words x' and v' such that for $w =_{\mathrm{def}} s_l(x_l) = s_l(x_l v_l)$ we have $x_l = x'w$ and $x_l v_l = v'w$ and it holds that $\delta(s_0, x') = s_3, \delta(s_3, w) = s_1, \delta(s_0, v') = s_4, \delta(s_4, w) = s_2, \delta(s_1, z_l) = s \in S'$ and $\delta(s_2, z_l) = s \notin S'$. We have to find simultaneous loops on the path w from s_3 to s_1 on one hand, and from s_4 to s_2 on the other hand to see that F has pattern $\mathbf{V}$. Therefore we look at the set of pairs $(p_i, q_i) \in S \times S$ for $0 \leq i \leq l$ with $p_i =_{\mathrm{def}} \delta(s_3, w_i)$ and $q_i =_{\mathrm{def}} \delta(s_4, w_i)$ where w_i is the i-prefix of w. Since there are only $|S|^2$ different such pairs there must be $i < j$ with $(p_i, q_i) = (p_j, q_j)$. Let $\tilde{w} \in A^+$ such that $w_j = w_i \tilde{w}$ and let w' such that $w = w_j w'$. Moreover there is some word $\tilde{v}$ with $\delta(p_i, \tilde{v}) = q_i$ since x' is a prefix of v'. Then we have pattern $\mathbf{V}$ witnessed by $\tilde{x} =_{\mathrm{def}} x' w_i, \tilde{w}, \tilde{v}$ and $\tilde{z} =_{\mathrm{def}} w' z_l$. ❑

With help of Statement 3 of Proposition 5.7 we may now reproof Theorem 3.11 using the previous proposition. We can even drop the minimality condition. Note that k can be bounded by $|S|^2$ which yields a polynomial time algorithm to determine the exact level k of a language in $\mathcal{B}_{1/2}$.

7 The Boolean Hierarchy over $\mathcal{B}_{1/2,k}$

Recall that $\mathcal{B}_{1,k} = \mathrm{BC}(\mathcal{B}_{1/2,k})$. We study the Boolean hierarchy over each class $\mathcal{B}_{1/2,k}$ to get an exact overview about the situation between $\mathcal{B}_{1/2,k}$ and $\mathcal{B}_{1,k}$. Especially, Lemma 2.1 applies to the Boolean hierarchy over each $\mathcal{B}_{1/2,k}$ since these classes are closed under union and intersection (see Proposition 5.6). So the Boolean hierarchy over $\mathcal{B}_{1/2,k}$ is equal to the family of classes $\mathcal{B}_{1/2,k}(l)$ and co$\mathcal{B}_{1/2,k}(l)$ for $l \geq 1$. We will obtain a membership criterion for these classes by examining the number of alternations that may occur in a sequence of words, where each word is a k–extension of its predecessor. From this we will obtain strictness and decidability results for the Boolean hierarchy over $\mathcal{B}_{1/2,k}$. Let us fix for this section some $k \geq 0$.

Definition 7.1. Let $L \subseteq A^+$, $m \geq 0$ and $w, v \in A^+$. We say that v is reachable from w by an m–alternating k–extension chain with respect to L, i.s. $w \overset{m,k}{\Longrightarrow}_L v$, if and only if there exist $w_0, \ldots, w_m \in A^+$ such that

1. $w = w_0 \preceq_k w_1 \preceq_k w_2 \preceq_k \ldots \preceq_k w_m \preceq_k v$, and
2. $w_i \in L$ if and only if $w_{i+1} \notin L$ for $0 \leq i \leq m-1$.

Next we take a closer look at such chains and define the sets of words that can be reached from a word (not) in a given language L by at least m alternations.

Definition 7.2. For a language $L \subseteq A^+$ and $m \geq 0$ we define

1. $L_k^+(m) =_{\mathrm{def}} \left\{ v \in A^+ \mid \exists w \, (w \in L \wedge w \overset{m,k}{\Longrightarrow}_L v) \right\}$.
2. $L_k^-(m) =_{\mathrm{def}} \left\{ v \in A^+ \mid \exists w \, (w \notin L \wedge w \overset{m,k}{\Longrightarrow}_L v) \right\}$.

We give useful properties of $L^+(m)$ and $L^-(m)$ in the following proposition.

Proposition 7.3. *For a language $L \subseteq A^+$ and $m \geq 0$ the following statements hold:*

1. $L_k^-(m) = \overline{L}_k^+(m)$.
2. $L_k^+(m+1) \cup L_k^-(m+1) \subseteq L_k^+(m) \cap L_k^-(m)$.
3. $L_k^+(m) = \langle L_k^+(m) \rangle_k$ *and* $L_k^-(m) = \langle L_k^-(m) \rangle_k$.
4. $L_k^+(m)$ *and* $L_k^-(m)$ *languages are in* $\mathcal{B}_{1/2,k}$.

Proof. To see Statement 1 suppose $v \in L_k^-(m)$ for some $m \geq 0$. If we look at the witnessing m–alternating k–extension chain then going from L to $\overline{L}$ just inverts its signature. Now this chain witnesses $v \in \overline{L}_k^+(m)$.

Statement 2 is due to the fact that an $m+1$–alternating k–extension chain is also an m–alternating k–extension chain. Statement 3 is immediate from the definitions. Statement 4 is an easy consequence of Statement 3 and Proposition 5.6. ❑

Now we show that any language L can be expressed as a possibly infinite union of set differences of sets $L_k^+(m)$ and $L_k^-(m)$.

Proposition 7.4. *For a language $L \subseteq A^+$ the following statements hold:*

1. $L = \bigcup_{m\geq 0}^{\infty} \left(L_k^+(2m) \backslash L_k^+(2m+1) \right)$ *and*
   $\overline{L} = \left(A^* \backslash L_k^+(0) \right) \cup \bigcup_{m\geq 1}^{\infty} \left(L_k^+(2m-1) \backslash L_k^+(2m) \right)$.
2. $\overline{L} = \bigcup_{m\geq 0}^{\infty} \left(L_k^-(2m) \backslash L_k^-(2m+1) \right)$ *and*
   $L = \left(A^* \backslash L_k^-(0) \right) \cup \bigcup_{m\geq 1}^{\infty} \left(L_k^-(2m-1) \backslash L_k^-(2m) \right)$.

Proof. It suffices to prove Statement 1. Statement 2 follows from 1 by Proposition 7.3.1. Let $m \geq 0$ and $v \in L_k^+(2m) \backslash L_k^+(2m+1)$. Because $v \in L_k^+(2m)$ there exists a $w \in L$ with $w \overset{2m,k}{\Longrightarrow}_L v$. Now observe that if $v \notin L$ then $w \overset{2m+1,k}{\Longrightarrow}_L v$ witnessed by the same k–extension chain as before, which is a contradiction to $v \notin L_k^+(2m+1)$. Hence $v \in L$.

Let $v \in L$ and look at all sequences of words $w_0, w_1, \ldots, w_l \in A^+$ such that $w_0 \in L$, $w_l = v$, $w_0 \preceq_k w_1 \preceq_k \ldots \preceq_k w_l$, and $w_i \neq w_{i+1}$ for $0 \leq i \leq l-1$. Note that such a sequence exists and in fact there is only a finite number of them, since $w_i \neq w_{i+1}$ and $w_i \preceq_k w_{i+1}$ implies that $|w_i| < |w_{i+1}|$. We can associate with each such sequence the number of alternations it has with respect to L. Let $l_{\max}$ be the maximal such number over all considered sequences. First observe that $l_{\max} = 2m$ for some $m \geq 0$ because $w_0, w_{l_{\max}} \in L$. Then we have $v \in L_k^+(2m)$ due to the sequence we chose $l_{\max}$ from. If $v \in L_k^+(2m+1)$ this contradicts the maximality of $l_{\max}$.

To see the statement for $\overline{L}$ one proves in the same way as above that $v \in L^+(2m-1) \backslash L^+(2m)$ implies $v \notin L$ for $m \geq 1$. Since for all $v \in A^+$ it holds that $v \overset{0,k}{\Longrightarrow}_L v$ we have $L \subseteq L_k^+(0)$. Hence $v \in A^* \backslash L^+(0)$ also implies $v \notin L$. On the other hand, if $v \notin L$ then there is no alternating k–extension chain starting with some $w \in L$ and ending with v, or we may argue as above to determine the maximal number of alternations which must be odd this time. ❑

In order to measure the number of inevitable alternations that occur with respect to a given language L we look for the maximal m such that the sets $L_k^+(m)$ and $L_k^-(m)$ are not empty.

Definition 7.5. For a language $L \subseteq A^+$ we set $m_k^+(L) =_{\mathrm{def}} \max\{\, m \mid L_k^+(m) \neq \emptyset \,\}$ and $m_k^-(L) =_{\mathrm{def}} \max\{\, m \mid L_k^-(m) \neq \emptyset \,\}$.

Since the measure m_k^+ gives the maximal number of alternations in any $\preceq_k$–chain, it is the same measure as used in Theorem 3.7. We relate the single classes of the Boolean hierarchy over $\mathcal{B}_{1/2,k}$ to particular values of m_k^+ and m_k^-. Our theorem then gives a membership criterion for the classes $\mathcal{B}_{1/2,k}(l)$ and has the equivalence of (1) and (3) in Theorem 3.7 as a corollary.

Theorem 7.6. *Let $L \subseteq A^+$ and let $l \geq 1$.*

1. *$L \in \mathcal{B}_{1/2,k}(l)$ if and only if $m_k^+(L) < l$.*
2. *$L \in \mathrm{co}\mathcal{B}_{1/2,k}(l)$ if and only if $m_k^-(L) < l$.*

Proof. We prove Statement 1 and we restrict ourselves to the case of even l, the other case being proved completely analogously.

Let $L \subseteq A^+$ with $m_k^+(L) < 2l$. Then $L_k^+(i) = \emptyset$ for all $i \geq 2l$. By Proposition 7.4.1 we can write L as

$$L = \bigcup_{i=0}^{l-1} \left(L_k^+(2i) \backslash L_k^+(2i+1) \right)$$

and Proposition 7.3, Statements 2 and 4 show that we can use Lemma 2.1.4 to obtain $L \in \mathcal{B}_{1/2,k}(2l)$.

Now suppose $L \in \mathcal{B}_{1/2,k}(2l)$. Then again by Lemma 2.1.4 there exist languages $L_1, L_2, \ldots, L_{2l} \in \mathcal{B}_{1/2,k}$ such that $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_{2l}$ and $L = \bigcup_{i=1}^{l} (L_{2i-1} \backslash L_{2i})$. Setting $L_0 =_{\mathrm{def}} A^+$ and $L_{2l+1} =_{\mathrm{def}} \emptyset$ we obtain $\overline{L} = \bigcup_{i=0}^{l} (L_{2i} \backslash L_{2i+1})$. Thus, each word from A^+ is contained in some set $L_i \backslash L_{i+1}$ for some $i \in \{0, \ldots, 2l\}$.

Assume that $L_k^+(2l) \neq \emptyset$. Then by definition of $L_k^+(2l)$ there exist $w \in L$, some $v \in L_k^+(2l)$ and $w_0, w_1, \ldots, w_{2l} \in A^+$ such that $w = w_0 \preceq_k w_1 \preceq_k w_2 \preceq_k \ldots \preceq_k w_{2l} \preceq_k v$ with $w_{2i} \in L$ and $w_{2i-1} \notin L$. For any $i \in \{0, 1, \ldots, 2l-1\}$ there must be two indices $j, j' \in \{0, \ldots, 2l\}$ with $w_i \in L_j \backslash L_{j+1}$ and $w_{i+1} \in L_{j'} \backslash L_{j'+1}$. Since $w_i \in L \Leftrightarrow w_{i+1} \notin L$ these indices must be different. Note with Proposition 5.6.1 that $\langle L_j \rangle_k = L_j$ for all j. So from $w_i \preceq_k w_{i+1}$ we can conclude that $w_{i+1} \in L_j$

as well, which implies $j' > j$. Consequently, the words $w_0, w_1, \ldots, w_{2l}$ are in $2l + 1$ different sets $L_j \backslash L_{j+1}$ with $j \geq 1$ (since $w_0 \in L \subseteq L_1$). This is a contradiction since there are only $2l$ such sets. Hence $m_k^+(L) < 2l$.

Statement 2 then follows from Statement 1 because $m_k^+(L) = m_k^-(\overline{L})$ which is immediate by Proposition 7.3.1. ❑

To see that the previous theorem yields the equivalence (1) ⇔ (3) of Theorem 3.7 recall from Lemma 2.1.1 that $\mathcal{B}_{1,k} = \bigcup_{l \geq 1} \mathcal{B}_{1/2,k}(l)$. We apply the provided criterion in two ways. First, we give a strictness argument for the Boolean hierarchy over $\mathcal{B}_{1/2,k}$ and then we show how the measures m_k^+ and m_k^- can be effectively computed.

Theorem 7.7. *For every $l \geq 1$ it holds that*

$$\mathcal{B}_{1/2,k}(l) \subsetneq \mathcal{B}_{1/2,k}(l+1).$$

Proof. Since $|A| \geq 2$ there are two different letters $a, b \in A$. Let $\alpha =_{\mathrm{def}} a^{k+1}$. For a word $w \in A^{\geq k+1}$ we define $|w|_\alpha$ to be the number of occurences of α in the k–decomposition $\widehat{w}$ of w. For $r \geq 1$ define

1. $M_{2r-1} =_{\mathrm{def}} \left\{ w \in A^{\geq k+1} \;\middle|\; |w|_\alpha \text{ is odd or } |w|_\alpha > 2r-1 \right\}$, and
2. $M_{2r} =_{\mathrm{def}} \left\{ w \in A^{\geq k+1} \;\middle|\; |w|_\alpha \text{ is odd and } |w|_\alpha \leq 2r \right\}$.

We claim that for all $l \geq 1$ it holds that $m_k^-(M_l) = l$ and $m_k^+(M_l) = l - 1$. Then by Theorem 7.6 we obtain $M_l \in \mathcal{B}_{1/2,k}(l) \backslash \mathrm{co}\mathcal{B}_{1/2,k}(l)$ and hence by Lemma 2.1.2 we get $\mathcal{B}_{1/2,k}(l) \subsetneq \mathcal{B}_{1/2,k}(l+1)$ for all $l \geq 1$. Thus it remains to prove our claim.

We first show that $m_k^-(M_l) \geq l$ and $m_k^+(M_l) \geq l - 1$ for all $l \geq 1$. Therefore, we define $y_i =_{\mathrm{def}} ba^k a^i$ for all $i \geq 0$. It is easy to see that $|y_i|_\alpha = i$ and that $y_i \preceq_k^{\mathrm{e}} y_{i+1}$ for all $i \geq 0$. As in the definition of the sets above we distinguish the cases for odd and even l, so let us assume $l = 2r-1$ for some $r \geq 1$. We see that $y_{2j} \notin M_{2r-1}$ and $y_{2j+1} \in M_{2r-1}$ for $0 \leq j \leq r-1$. So we may take $y_0 \preceq_k y_1 \preceq_k \ldots \preceq_k y_{2r-1}$ as an $2r-1$–alternating k–extension chain with respect to M_{2r-1} to witness that $m_k^-(M_{2r-1}) \geq 2r-1$. The same chain also witnesses that $m_k^+(M_{2r-1}) \geq 2r-2$.

Now let $l = 2r$ for some $r \geq 1$. Again, we observe that $y_{2j} \notin M_{2r}$ and $y_{2j+1} \in M_{2r}$ for $0 \leq j \leq r-1$. Since $y_{2r} \notin M_{2r}$ we may take $y_0 \preceq_k y_1 \preceq_k \ldots \preceq_k y_{2r}$ as an $2r$–alternating k–extension chain with respect to M_{2r} to witness that $m_k^-(M_{2r}) \geq 2r$. The same chain also witnesses that $m_k^+(M_{2r}) \geq 2r-1$.

We want to prove next that $m_k^-(M_l) \leq l$ and $m_k^+(M_l) \leq l-1$ for all $l \geq 1$. Suppose there is a k-extension chain involving words w_i. We note that from $w_i \preceq_k w_{i+1}$ it follows that $|w_i|_\alpha \leq |w_{i+1}|_\alpha$. Moreover, due to the definition of the sets M_l this inequality must be strict if there is an alternation between w_i and w_{i+1}.

First assume $l = 2r-1$ for some $r \geq 1$. If $m_k^-(M_{2r-1}) > 2r-1$ then there are $w_0, w_1, \ldots, w_{2r} \in A^+$ such that $w = w_0 \preceq_k w_1 \preceq_k w_2 \preceq_k \ldots \preceq_k w_{2r}$ with $w_{2i} \notin M_{2r-1}$ and $w_{2i+1} \in M_{2r-1}$. It follows by the previous remarks that $|w_{2r}|_\alpha > 2r-1$, a contradiction to $w_{2r} \notin M_{2r-1}$. If $m_k^+(M_{2r-1}) > 2r-2$ then there are $w_0, w_1, \ldots, w_{2r-1} \in A^+$ such that $w = w_0 \preceq_k w_1 \preceq_k w_2 \preceq_k \ldots \preceq_k w_{2r-1}$ with $w_{2i} \in M_{2r-1}$ and $w_{2i+1} \notin M_{2r-1}$. Since in particular $|w_0|_\alpha > 0$ it follows that $|w_{2r-1}|_\alpha > 2r-1$, a contradiction to $w_{2r-1} \notin M_{2r-1}$.

Now let $l = 2r$ for some $r \geq 1$. Suppose $m_k^-(M_{2r}) > 2r$. We may conclude as above that in a witnessing $(2r+1)$–alternating k–extension chain we have $|w_{2r+1}|_\alpha > 2r$, a contradiction to $w_{2r+1} \in M_{2r}$. If $m_k^+(M_{2r}) > 2r-1$ we obtain also as above for a witnessing chain that $|w_{2r}|_\alpha > 2r$ since $|w_0|_\alpha > 0$. This contradicts that $w_{2r} \in M_{2r}$. ❑

For a given dfa F, the equivalence $L(F) \in \mathcal{B}_{1/2,k}(l) \Leftrightarrow m_k^+(L(F)) < l$ given by Theorem 7.6 can be used to obtain a decision procedure for the question '$L(F) \in \mathcal{B}_{1/2,k}(l)$' for fixed l. In particular, we use that $m_k^+(L(F)) < l \Leftrightarrow L(F)_k^+(l) = \emptyset$. We will construct for $l \geq 1$ a nondeterministic finite automata F_l from F that accepts $L(F)_k^+(l)$. An emptyness test will then provide us with the answer to the above question. All this translates in an obvious way to the measure m_k^- (e.g., consider the dfa for $\overline{L}$). For nondeterministic finite automata (nfa) we let $\delta : A \times S \to 2^S$.

Lemma 7.8. *Let F be a dfa and $l \geq 1$. Then there exists an nfa F_l with $|F_l| \leq (2|F||A|^{2k+2})^{l+3}$ such that $L(F_l) = L(F)_k^+(l)$.*

Proof. Let $F = (A, S, \delta, s_0, S')$ and set $L =_{\mathrm{def}} L(F)$. We are interested in the set of words v that can be reached from some word $w \in L$ via a $\preceq_k$–chain having at least l alternations w.r.t. L. Note that from $l \geq 1$ it follows that $|v| > k+1$ and observe that $L \cap A^{\leq k+1} \subseteq L_k^+(0) \backslash L_k^+(1)$. The automaton F_l we have in mind guesses on input v a $\triangleleft_k$–chain of sufficient length and stores the k–prefix and the k–suffix of each word in the chain. It also remembers the states to which these words lead to in F. Then F_l accepts v if and only if we find l alternations with respect to S' in this sequence of states and if the stored k–prefixes and k–suffixes match the ones of the input.

For each word w_i of the $\triangleleft_k$–chain we store a quadruple $t_i = (c_i, d_i, f_i, r_i)$ where c_i (d_i) is the current k–prefix (k–suffix, resp.) of w_i, where f_i denotes a 0-1–valued variable, and r_i is the state F reaches after input w_i. The flag f_i tells us whether there is already a w_i in the guessed chain ($f_i = 1$) or whether the chain is still too short ($f_i = 0$). We also store in a variable c (d) the k–prefix (k–suffix, resp.) of the input. So let the set of states T be defined as the set of all tuples

$$[t_0, t_1, \ldots, t_l, c, d]$$

such that for $0 \leq i \leq l$ we have $t_i = (c_i, d_i, f_i, r_i)$ with $c_i, d_i \in A^{\leq k} \cup \{\epsilon\}$, $f_i \in \{0,1\}$, $r_i \in S$ and $c, d \in A^{\leq k} \cup \{\epsilon\}$. Moreover, we set

$$s_0^l =_{\mathrm{def}} [\underbrace{(\epsilon,\epsilon,0,s_0),(\epsilon,\epsilon,0,s_0),\ldots,(\epsilon,\epsilon,0,s_0)}_{(l+1)\text{–times}},\epsilon,\epsilon]$$

as the starting state. Let $[t_0, t_1, \ldots, t_l, c, d], [t'_0, t'_1, \ldots, t'_l, c', d'] \in T$ be given with $t'_i = (c'_i, d'_i, f'_i, r'_i)$ for $0 \leq i \leq l$. We define that $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l([t'_0, t'_1, \ldots, t'_l, c', d'], a)$ if and only if $c = p_k(c'a)$, $d = s_k(d'a)$ and there exists some j with $0 \leq j \leq l+1$ such that

1. for $0 \leq i < j$ it holds that
   (a) if $f'_i = 0$ then $t_i = (s_k(d'a), s_k(d'a), 0, \delta(s_0, s_k(d'a)))$,
   (b) if $f'_i = 1$ then $t_i = t'_i$,
2. and for $j \leq i \leq l$ it holds that
   (a) $d'_i = d'$ and $|d'_i| = k$, and
   (b) $t_i = (c'_i, s_k(d'a), 1, \delta(r'_i, a))$.

Let us comment a little bit on this definition before we continue. Clearly, on the next letter a of the input we want to maintain the actual k-prefix and k-suffix of the input in c and d. When reading the first k letters of the input there is no nondeterministic choice possible for j because of 2.(a), so it must be that $j = l+1$. Since during this time all f'_i are 0, only the variables for the k–prefixes and k–suffixes are filled and the respective states of F are stored (see 1.(a)). If the input is longer than k and if we have already guessed a chain $w_0 \triangleleft_k w_1 \triangleleft_k \ldots \triangleleft_k w_l \triangleleft_k v$, then we may guess a smallest index j such that the new last element $d'a$ of the input's k–decomposition appears in w_j first. We only obtain a $\triangleleft_k$–chain again if all words $w_j, w_{j+1}, \ldots, w_l, v$ have the same k–suffix, which is ensured by 2.(a). With 2.(b) we keep all old k–prefixes and set the new k–suffixes and states. The quadruples for the words $w_0, \ldots, w_{j-1}$ are not affected (see 1.(b)). It can also be the case that we have guessed only a chain $w_m \triangleleft_k \ldots \triangleleft_k w_l \triangleleft_k v$ with $m > 0$ yet. If $j \geq m$ we keep the actual values with 1.(a) and

1.(b). If $j < m$ then $w_j = w_{j+1} = \ldots = w_{m-1} = c'a = d'a$ are the new first words of the chain $w_j \lhd_k \ldots \lhd_k w_m \lhd_k \ldots \lhd_k w_l \lhd_k v$ and the flag variables $f_j, \ldots, f_{m-1}$ turn from 0 to 1 by 2.(b).

We want to prove formally what we just described. Note that due to the definition of δ_l only states $[t_0, t_1, \ldots, t_l, c, d]$ can be reached such that $f_0 f_1 \cdots f_l \in 0^*1^*$. Moreover, it holds that $|T| \leq (2|F||A|^{2k+2})^{l+3}$.

Claim. Let $v \in A^{\geq k+1}$. Then $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s_0^l, v)$ if and only if $c = p_k(v)$, $d = s_k(v)$ and there exist $w_m, w_{m+1}, \ldots, w_l \in A^{\geq k+1}$ for some m with $0 \leq m \leq l+1$ such that

1. $t_i = (d, d, 0, \delta(s_0, d))$ for $0 \leq i < m$,
2. $w_m \lhd_k w_{m+1} \lhd_k \ldots \lhd_k w_l \lhd_k v$ and
3. $t_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$ for $m \leq i \leq l$.

Proof of Claim. We prove this claim by induction on the length of v.

INDUCTION BASE. Let $v = xa$ with $x \in A^k, a \in A$ and set $s' =_{\text{def}} \delta(s_0, x)$. It is easy to see that $\delta_l(s_0^l, x) = \{[(x, x, 0, s'), \ldots, (x, x, 0, s'), x, x]\}$. Now if $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s_0^l, v)$ then there is some m with $0 \leq m \leq l+1$ such that $t_i = (c_i, d_i, 0, r_i)$ for $0 \leq i < m$ and $t_i = (c_i, d_i, 1, r_i)$ for $m \leq i \leq l$. Since $[t_0, t_1, \ldots, t_l, c, d]$ must emerge from $\delta_l(s_0^l, x)$ by choosing $j = m$ in the definition of δ_l, we have that $c = p_k(xa)$, $d = s_k(xa)$, that $t_i = (s_k(xa), s_k(xa), 0, \delta(s_0, s_k(xa)))$ for $0 \leq i < m$ and that $t_i = (x, s_k(xa), 1, \delta(s', a))$ for $m \leq i \leq l$. Clearly, since $xa = v$ we obtain $c = p_k(v)$, $d = s_k(v)$ and $t_i = (d, d, 0, \delta(s_0, d))$ for $0 \leq i < m$. If we define $w_i =_{\text{def}} v$ for $m \leq i \leq l$ then it holds that $w_m \lhd_k \ldots \lhd_k w_l \lhd_k v$. For $m \leq i \leq l$ we have $x = p_k(w_i)$, $s_k(xa) = s_k(w_i)$ and because $\delta(s', a) = \delta(\delta(s_0, x), a) = \delta(s_0, v) = \delta(s_0, w_i)$ we have shown the 'only-if'–part of the induction base.

To see the 'if'–part assume that $c = p_k(v)$, $d = s_k(v)$ and there are $w_m, \ldots, w_l \in A^{\geq k+1}$ for some m with $0 \leq m \leq l+1$ such that $w_m \lhd_k \ldots \lhd_k w_l \lhd_k v$. Then $c = p_k(xa)$, $d = s_k(xa)$ and we can conclude from $|v| = k+1$ that in fact $w_m = w_{m+1} = \ldots = w_l = v$. So the assumptions translate to $t_i = (s_k(xa), s_k(xa), 0, \delta(s_0, s_k(xa)))$ for $0 \leq i < m$ and $t_i = (p_k(xa), s_k(xa), 1, \delta(s_0, xa))$ for $m \leq i \leq l$. Now it can be seen that $[t_0, t_1, \ldots, t_l, c, d]$ is in $\delta_l(s_0^l, v)$ by choosing again $j = m$ in the definition of δ_l when going from $\delta_l(s_0^l, x)$ to $\delta_l(s_0^l, xa)$.

INDUCTION STEP. Let va be given with $v \in A^{\geq k+1}$ and $a \in A$. We want to show the 'only-if'–part first. Suppose we know that $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s_0^l, va)$ for some state $[t_0, t_1, \ldots, t_l, c, d] \in T$. Then there is some m with $0 \leq m \leq l+1$ such that $t_i = (c_i, d_i, 0, r_i)$ for $0 \leq i < m$ and $t_i = (c_i, d_i, 1, r_i)$ for $m \leq i \leq l$. Moreover, there is some $[t'_0, t'_1, \ldots, t'_l, c', d'] \in \delta_l(s_0^l, v)$ with $t'_i = (c'_i, d'_i, f'_i, r'_i)$ for $0 \leq i \leq l$ such that $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l([t'_0, t'_1, \ldots, t'_l, c', d'], a)$. Let j with $0 \leq j \leq l+1$ denote the nondeterministic choice of this transition.

By hypothesis we have $c' = p_k(v)$ and $d' = s_k(v)$, and by definition of δ_l it holds that $c = p_k(c'a)$ and $d = s_k(d'a)$. Hence $d = s_k(va)$ and because $|c'| = k$ we have $c = c' = p_k(v) = p_k(va)$. There also is some m' with $0 \leq m' \leq l+1$ such that $f'_i = 0$ for $0 \leq i < m'$ and $f'_i = 1$ for $m' \leq i \leq l$. Since $f_i = 0$ for $0 \leq i < m$ it must be that $m \leq m'$ and also $m \leq j$. So for $0 \leq i < m$ it holds by definition of δ_l that $t_i = (s_k(d'a), s_k(d'a), 0, \delta(s_0, s_k(d'a)))$. Hence $t_i = (d, d, 0, \delta(s_0, d))$ for $0 \leq i < m$ which shows Statement 1.

If $m < m'$ then it must be that $j = m$ because $f'_m = 0$ and $f_m = 1$. In this case the length of the guessed $\lhd_k$–chain strictly increases when going from v to va. If $m = m'$ then $m' \leq j \leq l$. We distinguish these two cases. Note that by hypothesis there exist $w'_{m'}, w'_{m'+1}, \ldots, w'_l \in A^{\geq k+1}$ such that $w'_{m'} \lhd_k w'_{m'+1} \lhd_k \ldots \lhd_k w'_l \lhd_k v$ and $t'_i = (p_k(w'_i), s_k(w'_i), 1, \delta(s_0, w'_i))$ for $m' \leq i \leq l$.

Case 1. Suppose $m < m'$ and hence $j = m$. So by definition of δ_l it holds that $d'_i = d'$ and $|d'_i| = k$ for $m \leq i \leq l$ and since $f'_m = f'_{m+1} = \ldots = f'_{m'-1} = 0$ we have by hypothesis that

$t'_i = (d', d', 0, \delta(s_0, d'))$ for $m \le i < m'$. Define for $m \le i < m'$ words $w_i =_{\rm def} d'a$ and for $m' \le i \le l$ set $w_i =_{\rm def} w'_i a$. For concise notations we set $w'_{l+1} =_{\rm def} v$ and $w_{l+1} =_{\rm def} va$. Then clearly, for $m \le i < m'-1$ we have $d'a = w_i \triangleleft_k w_{i+1} = d'a$. Moreover, for $m' \le i \le l$ it holds that $s_k(w'_i) = d'_i = d' = s_k(v)$ and hence we have $w_i \triangleleft_k w_{i+1}$ for $m' \le i \le l$. The latter is because $w'_i \triangleleft_k w'_{i+1}$ and we may additionally map the new last element $d'a$ of the k-decomposition of w_i to $s_{k+1}(w_{i+1}) = s_k(w'_{i+1})a$. For the same reason we have $w_{m'-1} = d'a = s_k(w'_{m'})a \triangleleft_k w'_{m'}a = w_{m'}$. Together we have obtained $w_m \triangleleft_k \ldots \triangleleft_k w_l \triangleleft_k va$ which gives Statement 2.

For $m \le i < m'$ we have by definition that $w_i = d'a$ and by hypothesis that $c'_i = d'$, $d'_i = d'$ and $r'_i = \delta(s_0, d')$. By definition of δ_l it holds that $t_i = (c'_i, s_k(d'a), 1, \delta(r'_i, a))$ and hence $t_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$. For $m' \le i \le l$ we have by definition $w_i = w'_i a$ and by hypothesis that $c'_i = p_k(w'_i)$, $d'_i = s_k(w'_i)$ and $r'_i = \delta(s_0, w'_i)$. By definition of δ_l we have $t_i = (c'_i, s_k(d'a), 1, \delta(r'_i, a))$ and since $d' = d'_i$ it follows that $t_i = (p_k(w'_i a), s_k(w'_i a), 1, \delta(s_0, w'_i a))$. Together we see that $t_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$ for $m \le i \le l$ which gives Statement 3.

Case 2. Now assume $m = m'$ and hence $m' \le j \le l$. Define for $m \le i < j$ words $w_i =_{\rm def} w'_i$ and for $j \le i \le l$ set $w_i =_{\rm def} w'_i a$. Again, set for concise notations $w'_{l+1} =_{\rm def} v$ and $w_{l+1} =_{\rm def} va$. Statements 2 and 3 are clear for w_i with $m \le i < j$ by hypothesis and since by definition of δ_l we have $t_i = t'_i$. From $w'_{j-1} \triangleleft_k w'_j$ we obtain $w_{j-1} = w'_{j-1} \triangleleft_k w'_j a = w_j$. By hypothesis we have $d'_i = s_k(w'_i)$ and due to the definition of δ_l it follows that $s_k(w'_i) = d'$ for $j \le i \le l$. So we can conclude as in Case 1 that $w_i = w'_i a \triangleleft_k w'_{i+1} a = w_{i+1}$ for $j \le i \le l$ and hence we have $w_m \triangleleft_k \ldots \triangleleft_k w_l \triangleleft_k va$.

Additionally to $d'_i = s_k(w'_i)$, it holds for $j \le i \le l$ by hypothesis that $c'_i = p_k(w'_i)$ and $r'_i = \delta(s_0, w'_i)$. By definition of δ_l we have as before for $j \le i \le l$ that $t_i = (c'_i, s_k(d'a), 1, \delta(r'_i, a))$ and it follows that $t_i = (p_k(w'_i a), s_k(w'_i a), 1, \delta(s_0, w'_i a)) = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$. This completes the proof for Statements 2 and 3 for Case 2.

Next we show the 'if'–part. Assume that $c = p_k(va)$, $d = s_k(va)$ and there exist $w_m, \ldots, w_l \in A^{\ge k+1}$ for some m with $0 \le m \le l+1$ such that $t_i = (d, d, 0, \delta(s_0, d))$ for $0 \le i < m$ and $w_m \triangleleft_k \ldots \triangleleft_k w_l \triangleleft_k va$ and $t_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$ for $m \le i \le l$. We need to show $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s_0^l, va)$ and start with the construction of a $\triangleleft_k$–chain for v from the given $\triangleleft_k$–chain for va in order to apply the hypothesis below.

Set $w_{l+1} =_{\rm def} va$, $w'_{l+1} =_{\rm def} v$, $c' =_{\rm def} p_k(v)$ and $d' =_{\rm def} s_k(v)$. We give an algorithm which computes a $\triangleleft_k$–chain for v stepwise for $n = l$ downto m. During this computation we will ensure that before each step n the following condition (C) holds.

(C) For $n+1 \le i \le l+1$ there are words w'_i such that $w_i = w'_i a$ and $s_k(w'_i) = d'$, and for $n+1 \le i \le l$ it holds that $w'_i \triangleleft_k w'_{i+1}$.

If we set initially $n = l$ then clearly (C) holds. We proceed as follows.

Step n:
(1) If $n < m$ then STOP.
(2) If $w_n \triangleleft_k w'_{n+1}$ then STOP.
(3) If $|w_n a^{-1}| = k$ then STOP.
(4) Set $w'_n =_{\rm def} w_n a^{-1}$ and continue with Step $n-1$.

Let us first argue, why before Step $n-1$ again (C) holds. After Line (1) we know that $n \ge m$. After Line (2) we have $w_n \not\triangleleft_k w'_{n+1}$. It follows that $s_{k+1}(w_n) = d'a$ since otherwise we obtain from $w_n \triangleleft_k w'_{n+1} a$ and $s_k(w'_{n+1}) = d'$ that $w_n \triangleleft_k w'_{n+1}$. With the definition in Line (4) we have $s_k(w'_n) = d'$ and since we know that $|w'_n| \ge k+1$ we obtain $w'_n \triangleleft_k w'_{n+1}$. The latter is because $w'_n a \triangleleft_k w'_{n+1} a$. It is clear that the proposed algorithm stops.

Case 1. Suppose our algorithm stops in Line (1) or Line (2). In the former case we have by (C) that for $m \le i \le l+1$ there are words w'_i such that $w_i = w'_i a$ and $s_k(w'_i) = d'$, and for $m \le i \le l$ it holds that $w'_i \triangleleft_k w'_{i+1}$. In the latter case we have $n \ge m$ and also by (C) that for $n+1 \le i \le l+1$ there are words w'_i such that $w_i = w'_i a$ and $s_k(w'_i) = d'$, and for $n+1 \le i \le l$ it holds that $w'_i \triangleleft_k w'_{i+1}$. If we define for $m \le i \le n$ words $w'_i =_{\rm def} w_i$ we obtain for $m \le i \le l$ that $w'_i \triangleleft_k w'_{i+1}$ since $w_n \triangleleft_k w'_{n+1}$.

Taking this together, we can state that for $m \le i \le l$ there are words w'_i such that $w'_i \lhd_k w'_{i+1}$. Moreover, for $m \le i \le n$ it holds that $w'_i = w_i$ and for $n+1 \le i \le l+1$ we have $s_k(w'_i) = d'$ and $w_i = w'_i a$. Note that $m = n+1$ is possible. By hypothesis we know that $[t'_0, t'_1, \ldots, t'_l, c', d'] \in \delta_l(s^l_0, v)$ with $t'_i = (d', d', 0, \delta(s_0, d'))$ for $0 \le i < m$ and $t'_i = (p_k(w'_i), s_k(w'_i), 1, \delta(s_0, w'_i))$ for $m \le i \le l$. Since for $n+1 \le i \le l+1$ we have $d'_i = s_k(w'_i) = d'$ we can apply δ_l to $[t'_0, t'_1, \ldots, t'_l, c', d']$ and input a with $j =_{\mathrm{def}} n+1$. Recall that by definition of δ_l we obtain a state $[t''_0, t''_1, \ldots, t''_l, c'', d'']$ with $c'' = p_k(c'a)$, $d'' = s_k(d'a)$ and

- $t''_i = (s_k(d'a), s_k(d'a), 0, \delta(s_0, s_k(d'a)))$ for $0 \le i < m$,
- $t''_i = t'_i$ for $m \le i \le n$ and
- $t''_i = (c'_i, s_k(d'a), 1, \delta(r'_i, a))$ for $n+1 \le i \le l$.

We have $c'' = c' = p_k(va) = c$ and $d'' = s_k(va) = d$. Moreover, it holds that $t''_i = (d, d, 0, \delta(s_0, d))$ for $0 \le i < m$ and we have $t''_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$ for $m \le i \le n$, because we have defined $w'_i = w_i$. Now let $n+1 \le i \le l$. Then $c'_i = p_k(w'_i) = p_k(w'_i a)$ and $d' = d'_i = s_k(w'_i)$ implies $s_k(d'a) = s_k(w'_i a)$. Finally, by $\delta(r'_i, a) = \delta(\delta(s_0, w'_i), a) = \delta(s_0, w'_i a)$ it follows that $t''_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$ for $n+1 \le i \le l$. Note that we have set $w'_i = w_i a^{-1}$. Together this shows $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s^l_0, va)$.

Case 2. Assume our algorithm stops in Line (3). Then $n \ge m$. Since it did not stop in Line (2) we have $s_k(w_n) = d'a$ and from $|w_n| = k+1$ it follows that $w_m = w_{m+1} = \ldots = w_n = d'a$. Moreover, by (C) we have for $n+1 \le i \le l+1$ that $w_i = w'_i a$ and $s_k(w'_i) = d'$, and for $n+1 \le i \le l$ that $w'_i \lhd_k w'_{i+1}$. By hypothesis we know that $[t'_0, t'_1, \ldots, t'_l, c', d'] \in \delta_l(s^l_0, v)$ with $t'_i = (d', d', 0, \delta(s_0, d'))$ for $0 \le i \le n$ and $t'_i = (p_k(w'_i), s_k(w'_i), 1, \delta(s_0, w'_i))$ for $n+1 \le i \le l$. Since we have for $m \le i \le n$ that $d'_i = d'$ and for $n+1 \le i \le l+1$ that $d'_i = s_k(w'_i) = d'$ we can apply δ_l to $[t'_0, t'_1, \ldots, t'_l, c', d']$ and input a with $j =_{\mathrm{def}} m$. Similar to above, we obtain by definition of δ_l a state $[t''_0, t''_1, \ldots, t''_l, c'', d'']$ with $c'' = p_k(c'a)$, $d'' = s_k(d'a)$ and

- $t''_i = (s_k(d'a), s_k(d'a), 0, \delta(s_0, s_k(d'a)))$ for $0 \le i < m$ and
- $t''_i = (c'_i, s_k(d'a), 1, \delta(r'_i, a))$ for $m \le i \le l$.

We conclude as in Case 1 that $c'' = c$, $d'' = d$ and that for $0 \le i < m$ and $n+1 \le i \le l$ we have $t''_i = t_i$. For $m \le i \le n$ we have $c''_i = c'_i = d' = p_k(d'a)$ and $\delta(r'_i, a) = \delta(\delta(s_0, d'), a) = \delta(s_0, d'a)$. Note that $w_i = d'a$ and hence $t''_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i))$ for $m \le i \le n$. This shows $[t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s^l_0, va)$ and finishes the proof of the claim. *(Proof of Claim.)* ❑

We want to specify the set of acccepting states for F_l. Let $[t_0, t_1, \ldots, t_l, c, d] \in T$ with $t_i = (c_i, d_i, f_i, r_i)$ for $0 \le i \le l$. We define $[t_0, t_1, \ldots, t_l, c, d] \in S'_l$ if and only if it holds for $0 \le i \le l$ that

- $f_i = 1$, $c_i = c$ and $d_i = d$, and
- $r_i \in S' \leftrightarrow i \equiv 0(2)$.

The first item ensures that we have guessed a chain $w_0 \preceq_k w_1 \preceq_k \ldots \preceq_k w_l \preceq_k v$, while the second one takes alternations with respect to the given dfa F into consideration. We set $F_l =_{\mathrm{def}} (A, T, \delta_l, s^l_0, S'_l)$ and claim that $L(F_l) = L(F)^+_k(l)$. To see this we conclude

$$
\begin{array}{rcl}
v \in L(F_l) & \Longleftrightarrow & \delta_l(s^l_0, v) \cap S'_l \neq \emptyset \\
& \Longleftrightarrow & \text{there exists some } [t_0, t_1, \ldots, t_l, c, d] \in \delta_l(s^l_0, v) \cap S'_l \\
& \Longleftrightarrow & c = p_k(v), d = s_k(v) \text{ and there exist } w_0, \ldots, w_l \text{ such that } w_0 \lhd_k \ldots \lhd_k w_l \lhd_k v \\
& & \text{and for } 0 \le i \le l \text{ it holds that } t_i = (p_k(w_i), s_k(w_i), 1, \delta(s_0, w_i)) \text{ with } p_k(w_i) = c, \\
& & s_k(w_i) = d \text{ and } (\delta(s_0, w_i) \in S' \leftrightarrow i \equiv 0(2)) \\
& \Longleftrightarrow & \text{there exist } w_0, \ldots, w_l \text{ such that } w_0 \preceq_k \ldots \preceq_k w_l \preceq_k v \text{ and} \\
& & (\delta(s_0, w_i) \in S' \leftrightarrow i \equiv 0(2)) \text{ for } 0 \le i \le l \\
& \Longleftrightarrow & v \in L(F)^+_k(l)
\end{array}
$$

where the third equivalence is due to our claim and the definition of S'_l. ❑

Lemma 7.9. *Let F be a dfa, $k \geq 0$ and $l \geq 1$. On input (F, k, l), the questions '$m_k^+(L(F)) < l$' and '$m_k^-(L(F)) < l$' are decidable in nondeterministic space $k \cdot l + l \cdot \log |F|$.*

Proof. As pointed out before, it holds that $m_k^+(L(F)) < l \Leftrightarrow L(F)_k^+(l) = \emptyset \Leftrightarrow L(F_l) = \emptyset$ where F_l is the nfa constructed in the proof of Lemma 7.8. Obviously, $L(F_l) = \emptyset$ is equivalent with the non-existence of a path between the starting state of F_l and one of its accepting states. Hence, we have to solve the graph non-accessibility problem for the transition graph of $F_l = (A, T, \delta_l, s_0^l, S'_l)$. To do this, we start with writing the initial state s_l on the tape and then continuesly guess a next input letter and a nondeterministic choice j for the transition function δ_l. We then determine with help of δ_l the next state of F_l, overwrite the actual state and check whether the new state is an accepting one. If so, the algorithm stops and accepts the input, otherwise we continue this procedure. The space needed to do this is dominated by the space needed to store a state of F_l, since the needed counters remain small. So this can be done in space $\leq k \cdot l + l \cdot \log |F|$. Since the respective complexity class is closed under complementation, this shows the lemma. ❑

Theorem 7.10. *For fixed $l \geq 1$ and $k \geq 0$, the membership problems for $\mathcal{B}_{1/2,k}(l)$ and co$\mathcal{B}_{1/2,k}(l)$ are in* NL.

So we are able to decide the question '$m_k^+(L(F)) < l$' for given dfa F and given $k \geq 0, l \geq 1$. With help of Statement 4 of Theorem 3.7 we can now bound l for fixed k, which yields an algorithm that computes $m_k^+(L(F))$ effectively. Hence, we can determine for a languages $L \in \mathcal{B}_{1,k}$ its exact location in the Boolean hierarchy over $\mathcal{B}_{1/2,k}$.

Proposition 7.11. *Let $k \geq 0$. There is a recursive function g_k which assigns $m_k^+(L(F))$ to a dfa F.*

Proof. To determine $m_k^+(L(F))$ for a given dfa F we may apply the NSPACE($k \cdot l + l \cdot \log |F|$)–algorithm given by Lemma 7.9 for $l = 1, 2, \ldots, 2^{2|A|^{k+2}(k+1)^2|F|} + 1$ answering the questions $m_k^+(L(F)) < l$. If the first positive answer occurs, then g_k outputs $l - 1$. If all answers are negative, then g_k outputs a special symbol reflecting the fact that $m_k^+(L(F)) = \infty$ which is due to Theorem 3.7. Note that this can also be done by a binary search, since the answer string is monotonic. ❑

The characterization of $\mathcal{B}_{1,k}$ such that a languages L is *not* in this class if and only if there exists a $\preceq_k$–chain with an infinite number of alternations with respect to L, leads in a natural way to a forbidden pattern characterization for this class. If there is an infinitely often alternating chain then we can identify an infinite sequence of insertion positions in elementary k–extensions that cause alternation. From this, one can obtain a sequence of positions which is strictly increasing, stricly decreasing or where insertion happens infinitely often in a factor inserted just before. In any of these three cases we can locate a number of positions where the same states must appear if a dfa for L gets the respective words as an input. These states give rise to a pattern in the transition graph of the dfa and a close examination of the occuring loops shows that these patterns are sufficient to derive an infinitely often alternating chain. This proof was carried out in [Sch99b] and it is substantially different from the one in [Ste85a]. From an analysis of this proof one can determine a number of alternations depending on the alphabet size, the automaton size and on k, which suffices to find the forbidden patterns. This leads to a self–contained proof of a theorem similar to Theorem 3.7, but involving different patterns.

8 The Boolean Hierarchy over $\mathcal{B}_{1/2}$

The goal of this section is to relate what we know about the the Boolean hierarchy over $\mathcal{B}_{1/2,k}$ to the Boolean hierarchy over $\mathcal{B}_{1/2}$. We already have by Proposition 3.5 that $\mathcal{B}_1 = \bigcup_{k \geq 0} \mathcal{B}_{1,k}$ and by Proposition 5.7 that $\mathcal{B}_{1/2} = \bigcup_{k \geq 0} \mathcal{B}_{1/2,k}$. This relation holds for every level of the respective Boolean hierarchies.

Proposition 8.1. *Let $l \geq 1$. Then it holds that*

1. $\mathcal{B}_{1/2,k}(l) \subseteq \mathcal{B}_{1/2,k+1}(l)$ *for all* $k \geq 0$ *and*
2. $\mathcal{B}_{1/2}(l) = \bigcup_{k \geq 0} \mathcal{B}_{1/2,k}(l)$.

Proof. To see Statement 1 let $L \in \mathcal{B}_{1/2,k}(l)$ which is witnessed by languages $L_1, L_2, \ldots, L_l \in \mathcal{B}_{1/2,k}$ as stated in Lemma 2.1. By Proposition 5.7 we have that $L_i \in \mathcal{B}_{1/2,k+1}$ for $1 \leq i \leq l$ and again by Lemma 2.1 that $L \in \mathcal{B}_{1/2,k+1}(l)$.

For Statement 2 we proceed similarly. Let $L \in \mathcal{B}_{1/2}(l)$ via languages $L_1, L_2, \ldots, L_l \in \mathcal{B}_{1/2}$. By Proposition 5.7 we have $L_i \in \mathcal{B}_{1/2,k_i}$ for $1 \leq i \leq l$ and suitable $k_i \geq 0$. Also by Proposition 5.7 we see that $L_i \in \mathcal{B}_{1/2,k}$ for $k =_{\rm def} \max_{1 \leq i \leq l} k_i$ and hence by Lemma 2.1 we obtain $L \in \mathcal{B}_{1/2,k}(l)$. Conversely, let $L \in \mathcal{B}_{1/2,k}(l)$. Since $\mathcal{B}_{1/2,k} \subseteq \mathcal{B}_{1/2}$ we immediately have $L \in \mathcal{B}_{1/2}(l)$. ❑

Of course, these relations also hold for the classes co$\mathcal{B}_{1/2,k}(l)$, which can be proved completely analogously. We show that the inclusion in Statement 1 is in fact strict.

Proposition 8.2. *Let $k \geq 0$ and $l \geq 1$. Then $\mathcal{B}_{1/2,k}(l) \subsetneq \mathcal{B}_{1/2,k+1}(l)$.*

Proof. For the languages $L_{k+1} = a^{k+1}A^+$ we have $L_{k+1} \in \mathcal{B}_{1/2,k+1} \subseteq \mathcal{B}_{1/2,k+1}(l)$, because

$$L_{k+1} = \bigcup_{w \in A^{k+1}} (a^{k+1} | A^{k+2} | w)_{k+1}.$$

To see that $L_{k+1} \notin \mathcal{B}_{1/2,k}(l)$ it suffices to argue that $L_{k+1} \notin \mathcal{B}_{1,k}$. Let $x_0 =_{\rm def} a^{k+2}$ and for $i \geq 0$ set $x_{2i+1} =_{\rm def} a^k b x_{2i}$ and $x_{2i+2} =_{\rm def} a x_{2i+1}$. Then for all $i \geq 0$ we have $x_i \preceq_k x_{i+1}$ and it holds that $x_{2i} \in L_{k+1}$ and $x_{2i+1} \notin L_{k+1}$. This gives rise to a $\preceq_k$–chain with an infinite number of alternations and hence $L_{k+1} \notin \mathcal{B}_{1,k}$ by Theorem 3.7. ❑

Clearly, if $u \preceq_{k+1} v$ then $u \preceq_k v$, and hence every $\preceq_{k+1}$–chain is also a $\preceq_k$–chain. It follows that $m^+_{k+1}(L) \leq m^+_k(L)$ and $m^+_{k+1}(L) \leq m^+_k(L)$ for all languages L and all $k \geq 0$.

Definition 8.3. Let $L \subseteq A^+$. We set $m^+(L) =_{\rm def} \min_{k \geq 0} m^+_k(L)$ and $m^-(L) =_{\rm def} \min_{k \geq 0} m^-_k(L)$.

The measures m^+ and m^- relate to the single classes of the Boolean hierarchy over $\mathcal{B}_{1/2}$.

Theorem 8.4. *Let $L \subseteq A^+$ and let $l \geq 1$.*

1. $L \in \mathcal{B}_{1/2}(l)$ *if and only if* $m^+(L) < l$.
2. $L \in$ co$\mathcal{B}_{1/2}(l)$ *if and only if* $m^-(L) < l$.

Proof. We only show Statement 1. By Proposition 8.1 we have that $L \in \mathcal{B}_{1/2}(l)$ if and only there exists some $k \geq 0$ such that $L \in \mathcal{B}_{1/2,k}(l)$. By Theorem 7.6 the latter holds if and only if there exists some $k \geq 0$ such that $m^+_k(L) < l$ which in turn is equivalent to $m^+(L) < l$ by the definition of m^+. ❑

In a straightforward way we obtain an upper bound for this measure.

Proposition 8.5. *Let F be a dfa and set $N =_{\rm def} |F|^3$. Then it holds that*

$$m^+(L(F)) < \infty \iff m^+(L(F)) < 2^{2|A|^{N+2}(N+1)^2|F|} + 1.$$

Proof. We only have to show the implication from left to right. Using Proposition 3.9 and Theorem 3.7 we obtain $m^+(L(F)) < \infty \iff L(F) \in \mathcal{B}_1 \iff m^+_N(L(F)) < 2^{2|A|^{N+2}(N+1)^2|F|} + 1$ which implies $m^+(L(F)) < 2^{2|A|^{N+2}(N+1)^2|F|} + 1$. ❑

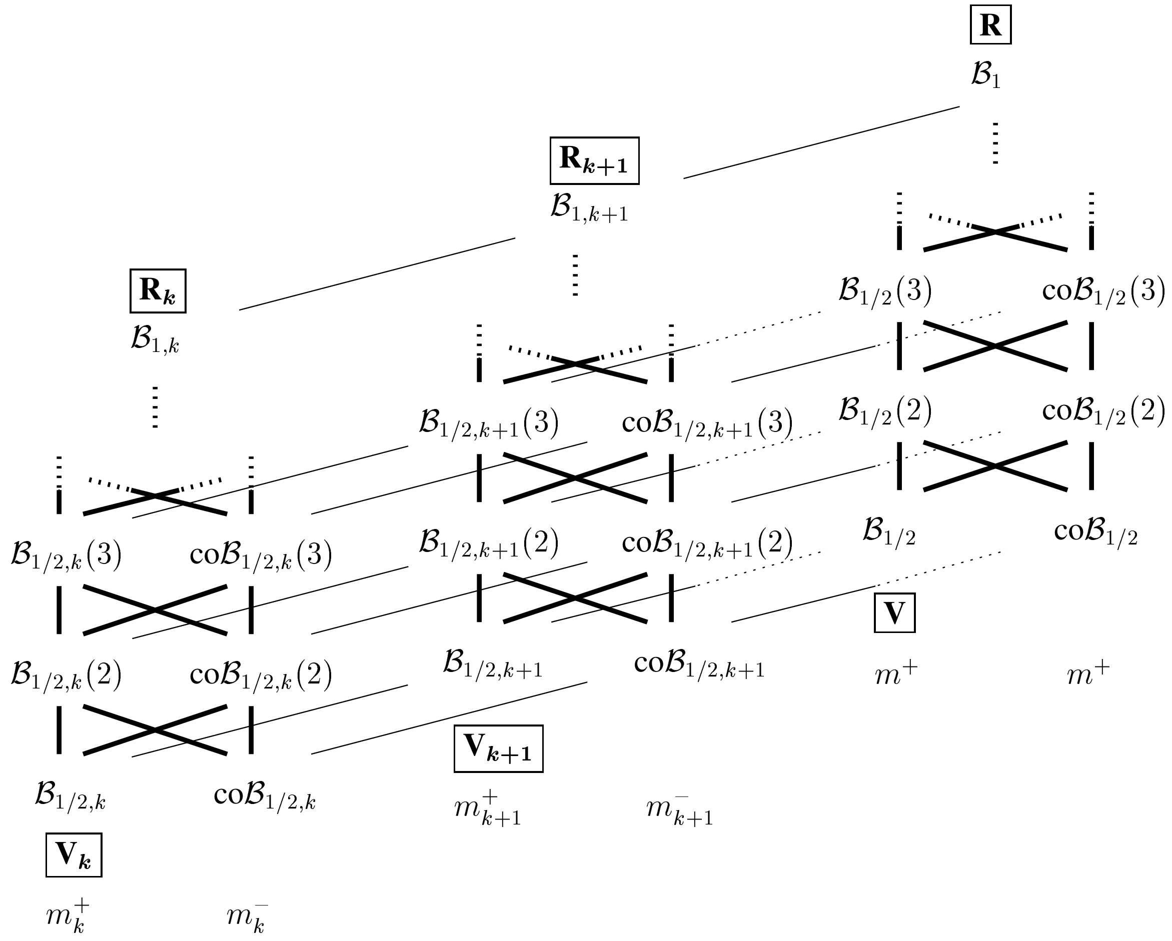

Figure 7: Inclusion structure of Boolean hierarchies inside dot–depth one. The membership problems for all classes are known to be decidable and all inclusions are strict.

In [Gla99] a characterization of the classes of the Boolean hierarchy over $\mathcal{B}_{1/2}$ was given in terms of a measure $m_{\mathcal{F}}$ for a given dfa F and hence $m^+(L(F)) = m_{\mathcal{F}}$ follows. It was shown there that the question '$m_{\mathcal{F}} < l$' is decidable and that the Boolean hierarchy over $\mathcal{B}_{1/2}$ is strict. In fact, for the latter the languages M_l for $k = 0$ from Theorem 7.7 can be used.

Let a dfa F be given accepting some language L. Taking together what we now know, we have complete control over the trade–off between the parameter k, which stands for the number of consecutive letters that needs to be spent for defining L with a fixed number of Boolean combination on one hand, and the number of inevitable Boolean combinations needed for fixed k on the other hand. This relation is given by the monotone decreasing mapping $k \mapsto m_k^+(L)$ which can be further analysed. For example, we can first minimize k such that $L \in \mathcal{B}_{1,k}$ and then compute $m_k^+(L)$, or we may first compute $m^+(L) = m_{\mathcal{F}}$ and then minimize k such that $m^+(L) = m_k^+(L)$.

9 Summary

We summarize the results of the structure of $\mathcal{B}_1$ in Figure 7. Recall that in case $k = 0$ we deal with the usual concatenation and hence we encounter the Boolean hierarchy over level $1/2$ of the Straubing–Thérien hierarchy. Generalizations of languages in $\mathcal{B}_{1,k}$ were studied in [Sch99a].

Acknowledgements. Thanks are due to Christian Glaßer for his comments and for reading proofs, and to Klaus W. Wagner, because parts of this paper are generalizations of methods developed in [SW98] for the case $k = 0$.

References

[Arf87] M. Arfi. Polynomial operations on rational languages. In *Proceedings 4th Symposium on Theoretical Aspects of Computer Science*, volume 247 of *Lecture Notes in Computer Science*, pages 198–206. Springer-Verlag, 1987.

[Arf91] M. Arfi. Opérations polynomiales et hiérarchies de concaténation. *Theoretical Computer Science*, 91:71–84, 1991.

[BK78] J. A. Brzozowski and R. Knast. The dot-depth hierarchy of star-free languages is infinite. *Journal of Computer and System Sciences*, 16:37–55, 1978.

[BKS98] B. Borchert, D. Kuske, and F. Stephan. On existentially first-order definable languages and their relation to NP. In *Proc. 25rd International Colloquium on Automata, Languages and Programming*, volume 1443 of *Lecture Notes in Computer Science*, pages 17–28. Springer Verlag, 1998.

[Brz76] J. A. Brzozowski. Hierarchies of aperiodic languages. *RAIRO Inform. Theor.*, 10:33–49, 1976.

[CB71] R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *Journal of Computer and System Sciences*, 5:1–16, 1971.

[CGH+88] J.-Y. Cai, T. Gundermann, J. Hartmanis, L. A. Hemachandra, V. Sewelson, K. W. Wagner, and G. Wechsung. The boolean hierarchy I: Structural properties. *SIAM Journal on Computing*, 17:1232–1252, 1988.

[CH91] S. Cho and D. T. Huynh. Finite-automaton aperiodicity is PSPACE-complete. *Theoretical Computer Science*, 88:99–116, 1991.

[CK96] C. Choffrut and J. Karhumäki. Combinatorics of words. In G.Rozenberg and A.Salomaa, editors, *Handbook of formal languages*, volume I, pages 329–438. Springer, 1996.

[Gla99] C. Glaßer. The Boolean hierarchy over dot-depth 1/2. Technical Report 230, Inst. für Informatik, Univ. Würzburg, 1999.

[Hig52] G. Higman. Ordering by divisibility in abstract algebras. In *Proc. London Math. Soc.*, volume 3, pages 326–336, 1952.

[Kna83] R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO Inform. Théor.*, 17:321–330, 1983.

[KSW87] J. Köbler, U. Schöning, and K. W. Wagner. The difference and the truth-table hierarchies for NP. *RAIRO Inform. Théor.*, 21:419–435, 1987.

[Pin96a] J. E. Pin. Logic, semigroups and automata on words. *Annals of Mathematics and Artificial Intelligence*, 16:343–384, 1996.

[Pin96b] J. E. Pin. Syntactic semigroups. In G.Rozenberg and A.Salomaa, editors, *Handbook of formal languages*, volume I, pages 679–746. Springer, 1996.

[PW97] J. E. Pin and P. Weil. Polynomial closure and unambiguous product. *Theory of computing systems*, 30:383–422, 1997.

[Sch99a] H. Schmitz. Generalized deterministic languages and their automata: A characterization of restricted temporal logic. Technical Report 226, Inst. für Informatik, Univ. Würzburg, 1999. An extended abstract appeared in the Preproceedings of DCAGRS'99, Magdeburg.

[Sch99b] H. Schmitz. Some forbidden patterns in automata for dot-depth one languages. Technical Report 220, Inst. für Informatik, Univ. Würzburg, 1999.

[Sim72] I. Simon. *Hierarchies of events with dot-depth one*. PhD thesis, University of Waterloo, 1972.

[SS83] J. Sakarovitch and I. Simon. Subwords. In M. Lothaire, editor, *Combinatorics on Words*, Encyclopedia of mathematics and its applications, pages 105–142. Addison-Wesley, 1983.

[Ste85a] J. Stern. Characterizations of some classes of regular events. *Theoretical Computer Science*, 35:17–42, 1985.

[Ste85b] J. Stern. Complexity of some problems from the theory of automata. *Information and Control*, 66:163–176, 1985.

[Str81] H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theoretical Computer Science*, 13:137–150, 1981.

[Str85] H. Straubing. Finite semigroups varieties of the form V * D. *J.Pure Appl.Algebra*, 36:53–94, 1985.

[SW98] H. Schmitz and K. W. Wagner. The Boolean hierarchy over level 1/2 of the Straubing-Thèrien hierarchy. Technical Report 201, Inst. für Informatik, Univ. Würzburg, 1998.

[Thé81] D. Thérien. Classification of finite monoids: the language approach. *Theoretical Computer Science*, 14:195–208, 1981.