The Boolean Hierarchy over Level $1/2$ of the Straubing-Thérien Hierarchy

Heinz Schmitz   Klaus W. Wagner
Theoretische Informatik
Universität Würzburg
Am Exerzierplatz 3
97072 Würzburg
{schmitz, wagner}@informatik.uni-wuerzburg.de

Abstract

For some fixed alphabet A with $|A| \geq 2$, a language $L \subseteq A^*$ is in the class $\mathcal{L}_{1/2}$ of the Straubing-Thérien hierarchy if and only if it can be expressed as a finite union of languages $A^*a_1A^*a_2A^* \cdots A^*a_nA^*$, where $a_i \in A$ and $n \geq 0$. The class $\mathcal{L}_1$ is defined as the boolean closure of $\mathcal{L}_{1/2}$. It is known that the classes $\mathcal{L}_{1/2}$ and $\mathcal{L}_1$ are decidable. We give a membership criterion for the single classes of the boolean hierarchy over $\mathcal{L}_{1/2}$. From this criterion we can conclude that this boolean hierarchy is proper and that its classes are decidable. In finite model theory the latter implies the decidability of the classes of the boolean hierarchy over the class Σ_1 of the FO[<]-logic. Moreover we prove a "forbidden-pattern" characterization of $\mathcal{L}_1$ of the type: $L \in \mathcal{L}_1$ if and only if a certain pattern does not appear in the transition graph of a deterministic finite automaton accepting L. We discuss complexity theoretical consequences of our results.

Classification: finite automata, concatenation hierarchies, boolean hierarchy, decidability

1 Introduction

We contribute to the theory of finite automata and regular languages, as well as to complexity theory. Particularly we deal with starfree regular languages. These are languages which are constructed from alphabet letters only by using boolean operations together with concatenation. Alternating these two kinds of operations in order to distinguish between combinatorial and sequential aspects leads to the definition of concatenation hierarchies that exhaust the class of starfree languages.

Prominent examples are the dot-depth hierarchy, first studied in [CB71], and the Straubing-Thérien hierarchy [Str81, Thé81, Str85]. Both are known to be strict [BK78] and closely related to each other. Most naturally arising questions concerning these hierarchies are of major interest in different research areas since there are close connections to finite model theory, theory of finite semigroups, topology, boolean circuits and others. For an overview or as a good starting point to this rich field of research see e.g. the articles [Brz76, Pin96a, Pin96b, Tho96].

In this paper we deal with the so-called Straubing-Thérien hierarchy. Let A be some finite alphabet with $|A| \geq 2$. For a class $\mathcal{C}$ of languages over A^* let $\mathrm{POL}(\mathcal{C})$ be its polynomial closure, i.e. the class of languages L that can be written as a finite union of languages $L_0a_1L_1a_2L_2\cdots L_{n-1}a_nL_n$, where $a_i \in A$, $L_i \in \mathcal{C}$ and $n \geq 0$. Denote by $\mathrm{BC}(\mathcal{C})$ its boolean closure, i.e. the closure of $\mathcal{C}$ under finite union, finite intersection and complementation. Then the Straubing-Thérien hierarchy can be defined as the family of classes $\mathcal{L}_{n/2}$, where we define $\mathcal{L}_0 =_{\mathrm{def}} \{\emptyset, A^*\}$, $\mathcal{L}_{n+1/2} =_{\mathrm{def}} \mathrm{POL}(\mathcal{L}_n)$, and $\mathcal{L}_{n+1} =_{\mathrm{def}} \mathrm{BC}(\mathcal{L}_{n+1/2})$ for $n \geq 0$ (notations are adopted from [PW97]). We will also consider the classes $\mathrm{co}\mathcal{L}_{n+1/2}$, where $\mathrm{co}\mathcal{C} =_{\mathrm{def}} \{\, \overline{L} \mid L \in \mathcal{C} \,\}$ for a class $\mathcal{C}$. It was shown by M. Arfi in [Arf87, Arf91] that the classes $\mathcal{L}_{n+1/2}$ (and $\mathrm{co}\mathcal{L}_{n+1/2}$) are closed under intersection. For a language $L \subseteq A^*$ and a minimal n with $L \in \mathcal{L}_{n/2}$ we say that L *has level* $n/2$.

The connection between first-order logic and the class of starfree languages goes back to the work of McNaughton and Papert [MP71]. The Straubing-Thérien hierarchy is related to the first-order logic $\mathrm{FO}[<]$ having only the binary relation $<$ and unary relations for the alphabet symbols from A. Let Σ_k be the subclass of $\mathrm{FO}[<]$ which is defined by at most $k-1$ quantifier alternations, starting with an existential quantifier. It has been proved by W. Thomas in [Tho82] (see also [PP86]) that Σ_k formulas describe just the $\mathcal{L}_{k-1/2}$ languages and that the boolean combinations of Σ_k formulas describe just the $\mathcal{L}_k$ languages.

Unfortunately one main question about the Straubing-Thérien hierarchy, namely the question of the decidability of its classes, appears to be extremely difficult, although a lot of effort via different approaches has been invested. The decidability problem can be stated as follows: given some $n \geq 0$ and a regular language L presented by a deterministic finite automaton, decide whether or not L has level $n/2$. To our knowledge, only levels $0, 1/2, 1$, and $3/2$ are known to be decidable (cf. [PW97]).

The purpose of this paper is to start with an exact analysis of what happens between level $1/2$ and level 1. Since $\mathcal{L}_1 = \mathrm{BC}(\mathcal{L}_{1/2})$ and since $\mathrm{BC}(\mathcal{L}_{1/2})$ is just the union of the classes $\mathcal{L}_{1/2}(k)$ of the boolean hierarchy over $\mathcal{L}_{1/2}$ we study these classes $\mathcal{L}_{1/2}(k)$ and their decidability.

J. Stern [Ste85] proved the following interesting characterization of the class $\mathcal{L}_1$ (the class of piecewise testable languages over alphabet A): A language $L \subseteq A^*$ is in $\mathcal{L}_1$ if and only if there does not exist an infinite chain $w_1, w_2, w_3, \ldots$ of words where w_{i+1} is an extension of w_i and $w_i \in L \Leftrightarrow w_{i+1} \notin L$ for $i = 1, 2, 3, \ldots$. Let $m^+(L)$ be the length of a maximal chain of this kind starting with $w_1 \in L$. Using a normal form theorem for classes of boolean hierarchies, we prove that $L \in \mathcal{L}_{1/2}(k)$ if and only if $m^+(L) < k$. Since the latter property can be decided for fixed k with a nondeterministic logarithmic space algorithm, we can also decide the membership problem for the classes $\mathcal{L}_{1/2}(k)$ with a nondeterministic logarithmic space algorithm. Furthermore we show that the measure $m^+(L)$ is computable with an exponential space algorithm. Another consequence of the above membership criterion for the classes $\mathcal{L}_{1/2}(k)$ is the fact that this boolean hierarchy is indeed proper.

As a second contribution we prove a "forbidden-pattern" characterization of $\mathcal{L}_1$ of the type: $L \in \mathcal{L}_1$ if and only if a certain pattern (see Figure 3) does not appear in a deterministic finite automaton accepting L. Such characterizations were already known for the classes $\mathcal{L}_{1/2}$ and $\mathcal{L}_{3/2}$ [PW97]. Our characterization easily provides a nondeterministic logspace decision algorithm for $\mathcal{L}_1$.

There is a close connection between concatenation hierarchies and complexity classes, both related via the so-called leaf language approach to define complexity classes. This approach was introduced in [BCS92, Ver93] and led to a number of interesting results (cf. [HLS$^+$93, JMT94, BV98, CHVW98]). In particular in [BV98] it was shown that taking the languages from $\mathcal{L}_{k-1/2}$ as leaf languages yields exactly the k-th class of the polynomial time hierarchy. In the last section we state a result of this type relating the boolean hierarchy over level $1/2$ of the Straubing-Thérien hierarchy to the boolean hierarchy over NP. A similar, but ineffective result concerning the boolean hierarchy over level $1/2$ of the dot-depth hierarchy was obtained in [BKS98]. Here we can make use of our decision algorithm, which is not known for the case of the dot-depth hierarchy.

Finally we want to make a remark concerning our methods. First we note that the normalform results we use for the classes of the boolean hierarchy over $\mathcal{L}_{1/2}$ are valid also for the classes of the boolean hierarchy over every class $\mathcal{L}_{n+1/2}$. This combined with the "forbidden-pattern" technique could work to achieve similar structural and decidability results for every level of the Straubing-Thérien hierarchy.

2 Preliminaries

We consider languages over an arbitrary finite alphabet A with $|A| \geq 2$. For a class $\mathcal{C}$ of languages, 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. 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

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

where $\mathcal{C} \oplus \mathcal{C} =_{\mathrm{def}} \{A \triangle B \mid A \in \mathcal{C},\ B \in \mathcal{C}\}$, $\triangle$ denotes the symmetric set difference and $\mathrm{co}\mathcal{C} =_{\mathrm{def}} \left\{ \overline{L} \mid L \in \mathcal{C} \right\}$.

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 $k \geq 1$.*

1. $\mathrm{BC}(\mathcal{C}) = \bigcup_{k \geq 1} \mathcal{C}(k)$.
2. $\mathcal{C}(k) \cup \mathrm{co}\mathcal{C}(k) \subseteq \mathcal{C}(k+1) \cap \mathrm{co}\mathcal{C}(k+1)$.
3. $L \in \mathcal{C}(2k-1)$ *if and only if there exist languages* $L_1, L_2, \ldots, L_{2k-1} \in \mathcal{C}$ *such that* $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_{2k-1}$ *and* $L = \bigcup_{i=1}^{k-1}(L_{2i-1} \backslash L_{2i}) \cup L_{2k-1}$.
4. $L \in \mathcal{C}(2k)$ *if and only if there exist languages* $L_1, L_2, \ldots, L_{2k} \in \mathcal{C}$ *such that* $L_1 \supseteq L_2 \supseteq \cdots \supseteq L_{2k}$ *and* $L = \bigcup_{i=1}^{k}(L_{2i-1} \backslash L_{2i})$.

For a class $\mathcal{C}$ of languages, let $\mathrm{POL}(\mathcal{C})$ be its polynomial closure, i.e. the class of languages L that can be written as a finite union of languages $L_0a_1L_1a_2L_2\cdots L_{n-1}a_nL_n$, where $a_i \in A$, $L_i \in \mathcal{C}$ and $n \geq 0$. Then the Straubing-Thérien hierarchy can be defined as the following family of classes, where notations are adopted from [PW97].

1. $\mathcal{L}_0 =_{\mathrm{def}} \{\emptyset, A^*\}$
2. $\mathcal{L}_{n+1/2} =_{\mathrm{def}} \mathrm{POL}(\mathcal{L}_n)$ for $n \geq 0$
3. $\mathcal{L}_{n+1} =_{\mathrm{def}} \mathrm{BC}(\mathcal{L}_{n+1/2})$ for $n \geq 0$

We will also take into consideration the classes co$\mathcal{L}_{n+1/2}$. Any class $\mathcal{L}_{n+1/2}$ can be equivalently defined as the closure of the class $\mathcal{L}_n$ under union, intersection and the so-called marked concatenation (cf. [Arf87, Arf91]). Consequently, the results of Lemma 2.1 apply also to the classes $\mathcal{C} = \mathcal{L}_{n+1/2}$. For a language $L \subseteq A^*$ and a minimal n with $L \in \mathcal{L}_{n/2}$ we say that L *has level* $n/2$.

Next we point out a very natural connection between the Straubing-Thérien hierarchy and a certain logic over finite words. We define formulas using the binary relation symbol $<$ and unary relation symbols π_a for each letter $a \in A$. Atomic formulas are of the type $x < y$, $x = y$ and $\pi_a x$, with variables x, y. Then formulas are contructed from atomic formulas by using the connectives $\neg, \vee, \wedge$ and quatifiers $\exists, \forall$ bounding variables. Let Σ_k (Π_k) be the subclass of such formulas which have at most $k-1$ quantifier alternations, starting with an existential (universal, resp.) quantifier. We say a language $L \subseteq A^*$ is FO[<]-definable if there exists a sentence ϕ (i.e. a formula of the above type without free variables) such that all words $w \in L$ satisfy ϕ when variables are interpreted as positions in w, $\pi_a x$ means the letter at position x is a, and $<$ is the usual $<$-relation on $\{1, \ldots, |w|\}$.

Theorem 2.2 [Tho82, PP86]. *Let $k \geq 1$, and let $L \subseteq A^*$ be any language.*

1. *$L \in \mathcal{L}_{k-1/2}$ if and only if L is* FO[<]*-definable by a Σ_k formula.*
2. *$L \in$ co$\mathcal{L}_{k-1/2}$ if and only if L is* FO[<]*-definable by a Π_k formula.*
3. *$L \in \mathcal{L}_k$ if and only if L is* FO[<]*-definable by a boolean combination of Σ_k formulas.*

Let ϵ be the empty word. We denote by $\preceq$ the subword relation on A^*, i.e. $w \preceq v$ if and only if there exist $n \geq 1$, $a_1, a_2, \ldots, a_n \in A$ and $v_0, v_1, \ldots, v_n \in A^*$ such that $w = a_1a_2\cdots a_n$ and $v = v_0a_1v_1a_2v_2\cdots a_nv_n$. For $w \in A^*$ we define $\langle w \rangle_{\preceq} =_{\mathrm{def}} \{v \mid w \preceq v\}$ as the set of all words having w as a subword, i.e. $\langle a_1a_2\cdots a_n\rangle_{\preceq} = A^*a_1A^*a_2A^*\cdots A^*a_nA^*$ for all $n \geq 1$ and $a_1, a_2, \ldots, a_n \in A$. Moreover, for a language L let $\langle L \rangle_{\preceq} =_{\mathrm{def}} \bigcup_{w \in L}\langle w \rangle_{\preceq}$ be the set of all words having a subword in L. For a word $w = a_1a_2\cdots a_n$ we denote with w^R its reverse, i.e. $w^R =_{\mathrm{def}} a_na_{n-1}\cdots a_1$, and for a language L let $L^R =_{\mathrm{def}} \left\{ w^R \;\middle|\; w \in L \right\}$. We will denote infinite sequences of words $\{w_i\}_{i=1}^{\infty}$ for short as $\{w_i\}$.

As is standard, 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 set of states, $\delta : A \times S \to S$ is its transition function, $s_0 \in S$ is the starting state and $S' \subseteq S$ is the set of accepting states. We consider nondeterministic finite automata (nfa) as well, where $\delta : A \times S \to 2^S$. With $L(F)$ we denote the language accepted by an automaton F. As usual we extend transition functions to input words, and we denote by $|F|$ the number of states of F.

Theorem 2.3. *For every $L \subseteq A^*$ the following are equivalent:*

(1) $L \in \mathcal{L}_{1/2}$

(2) L is a finite union of sets $\langle w \rangle_{\preceq}$

(3) L is regular and $\langle L \rangle_{\preceq} = L$

Proof. The equivalence (1) ⇔ (2) is by definition, and (2) ⇒ (3) is obvious. For (3) ⇒ (2), let F be a dfa such that $\langle L(F) \rangle_{\preceq} = L(F)$. Let F' be the nfa which is constructed from F be introducing for every state and every $a \in A$ a simple loop with a. Obviously, $L(F') = L(F)$. Now convert F' into the nfa F'' by removing all nontrivial loops, i.e. by keeping only the paths leading directly from the starting state to an accepting state. Also, $L(F'') = L(F')$. Now, $L(F'')$ is the union of all $\langle a_1 a_2 \cdots a_n \rangle_{\preceq}$ where $a_1 a_2 \cdots a_n$ is a path in F'' leading directly from the starting state to an accepting state. ❑

We assume the reader to be familiar with complexity classes of common interest such as NL, P, NP and the levels Σ_k^{p} of the polynomial time hierarchy.

3 Alternating Word Extension Chains

We will obtain a membership criterion for the classes $\mathcal{L}_{1/2}(k)$ by examining the number of alternations that may occur in a sequence of words, where each word is an extension of its predecessor. Let us first make this notion precise.

Definition 3.1 [Ste85]. Let $L \subseteq A^*$, $m \geq 0$ and $w, v \in A^*$. We say that v *is reachable from w by an m-alternating word extension chain with respect to L*, i.s. $w \overset{m}{\Longrightarrow}_L v$, if and only if there exist $w_0, w_1, \ldots, w_m \in A^*$ such that

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

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 3.2. For a language $L \subseteq A^*$ and $m \geq 0$ we define

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

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

Proposition 3.3. *For a language L and $m \geq 0$ the following statements hold:*

1. $L^-(m) = \overline{L}^+(m)$.

2. $L^+(0) = \langle L \rangle_{\preceq}$ *and* $L^-(0) = \langle \overline{L} \rangle_{\preceq}$.

3. $L^+(m+1) \cup L^-(m+1) \subseteq L^+(m) \cap L^-(m)$.

4. $v \notin L^+(m) \cup L^-(m)$ *for all* $m > |v|$.

5. $\bigcap_{m \geq 0} L^+(m) = \bigcap_{m \geq 0} L^-(m) = \emptyset$.

6. $L^+(m) \neq \emptyset$ *implies* $L^+(m+1) \subsetneq L^+(m)$, *and* $L^-(m) \neq \emptyset$ *implies* $L^-(m+1) \subsetneq L^-(m)$.

7. $L^+(m) = \langle L^+(m) \rangle_{\preceq}$ *and* $L^-(m) = \langle L^-(m) \rangle_{\preceq}$.

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

Proposition 3.4. *For a language $L \subseteq A^*$ the following statements hold:*

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

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

Proof. Let $m \geq 0$ and $v \in L^+(2m) \backslash L^+(2m+1)$. Because of $v \in L^+(2m)$ there exists a $w \in L$ with $w \overset{2m}{\Longrightarrow}_L v$. Now observe that if $v \notin L$ then $w \overset{2m+1}{\Longrightarrow}_L v$ witnessed by the same word extension chain as before, which is a contradiction to $v \notin L^+(2m+1)$. Hence $v \in L$.

In the same way one proves that $v \in L^+(2m-1) \backslash L^+(2m)$ implies $v \notin L$ for $m \geq 1$, and that $v \in A^* \backslash L^+(0)$ implies $v \notin L$.

Statement 2 follows from 1 by Proposition 3.3.1. ❑

Now we want to show that for a regular set L the sets $L^+(m)$ and $L^-(m)$ belong to $\mathcal{L}_{1/2}$. Proposition 3.3.7 already says that $L^+(m) = \langle L^+(m) \rangle_{\preceq}$ and $L^-(m) = \langle L^-(m) \rangle_{\preceq}$. With Theorem 2.3 it remains to show that they are regular.

Lemma 3.5. *If $L \subseteq A^*$ is regular and $m \geq 0$, then $L^+(m)$ and $L^-(m)$ are regular as well.*

Proof. Let $F = (A, S, \delta, s_0, S')$ be a deterministic finite automaton accepting L. We construct a nondeterministic finite automaton F_m that accepts $L^+(m)$ and that realizes the idea of guessing a m-alternating chain of subwords of the input. Define $F_m =_{\mathrm{def}} (A, S_m, \delta_m, s_0^m, S'_m)$ as

- $S_m =_{\mathrm{def}} \underbrace{S \times S \times \cdots \times S}_{(m+1)-\mathrm{times}}$

- $s_0^m =_{\mathrm{def}} (s_0, s_0, \ldots, s_0)$

- $\delta_m((s_1, s_2, \ldots, s_{m+1}), a) =_{\mathrm{def}}$
  $$\big\{ (s_1, s_2, \ldots, s_i, \delta(s_{i+1}, a), \ldots, \delta(s_{m+1}, a)) \bigm| 0 \leq i \leq m+1 \big\}$$

- $S'_m =_{\mathrm{def}} \{ (s_1, s_2, \dots, s_{m+1}) \mid (s_i \in S' \leftrightarrow i \text{ odd}) \text{ for } i = 1, \dots, m+1 \}$.

We observe that $(s_1, s_2, \dots, s_{m+1}) \in \delta_m(s_0^m, v)$ if and only if there exist words $w_1, \dots, w_{m+1} \in A^*$ such that $w_1 \preceq w_2 \preceq \dots \preceq w_{m+1} \preceq v$ and $\delta(s_0, w_i) = s_i$ for $i = 1, \dots, m+1$.

Now we can conclude:

$$\begin{aligned} v \in L(F_m) &\iff \delta_m(s_0^m, v) \cap S'_m \neq \emptyset \\ &\iff \text{there exist } s_1, s_2, \dots, s_{m+1} \text{ such that } (s_1, s_2, \dots, s_{m+1}) \in \delta_m(s_0^m, v) \cap S'_m \\ &\iff \text{there exist } w_1, \dots, w_{m+1} \text{ such that } w_1 \preceq w_2 \preceq \dots \preceq w_{m+1} \preceq v \text{ and} \\ &\qquad (\delta(s_0, w_i) \in S' \leftrightarrow i \text{ odd}) \text{ for } i = 1, \dots, m+1 \\ &\iff v \in L^+(m) \end{aligned}$$

Because of $L^-(m) = \overline{L}^+(m)$ we obtain that $L^-(m)$ is also regular. ❑

Corollary 3.6. *If $L \subseteq A^*$ is regular and $m \geq 0$, then $L^+(m)$ and $L^-(m)$ are in $\mathcal{L}_{1/2}$.*

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 $L^+(m)$ and $L^-(m)$ are not empty.

Definition 3.7. For a language $L \subseteq A^*$ we set $m^+(L) =_{\mathrm{def}} \max\{ m \mid L^+(m) \neq \emptyset \}$ and $m^-(L) =_{\mathrm{def}} \max\{ m \mid L^-(m) \neq \emptyset \}$.

The following proposition is an immediate consequence of Proposition 3.3.

Proposition 3.8. *For any language $L \subseteq A^*$ it holds that*

1. *$m^+(L) = \infty$ if and only if $m^-(L) = \infty$,*
2. *if $m^+(L) < \infty$ then $|m^+(L) - m^-(L)| = 1$, and*
3. *$m^+(L) = m^-(\overline{L})$.*

4 A Criterion for Membership in $\mathcal{L}_{1/2}(k)$

The measure m^+ has already been used by J. Stern to characterize $\mathcal{L}_1 = \mathrm{BC}(\mathcal{L}_{1/2})$, i.e. the piecewise testable languages over alphabet A.

Theorem 4.1 [Ste85]. *A language $L \subseteq A^*$ belongs to $\mathcal{L}_1$ if and only if $m^+(L)$ is finite.*

Now we will relate the single classes of the boolean hierarchy over $\mathcal{L}_{1/2}$ to particular values of m^+ and m^-. This theorem then has the preceding one as a corollary.

Theorem 4.2. *Let $L \subseteq A^*$ be a language and $k \geq 1$.*

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

Proof. We prove Statement 1; Statement 2 then follows immediately by Proposition 3.8.3. We restrict ourselves to the case of even k, the other case being proved analogously.

Let L be regular and $m^+(L) < 2k$. Then $L^+(i) = \emptyset$ for all $i \geq 2k$. By Proposition 3.4.1 we can write L as

$$L = \bigcup_{i=0}^{k-1} \big(L^-(2i) \backslash L^-(2i+1)\big),$$

and Corollary 3.6 shows that we can use Lemma 2.1.4 to obtain $L \in \mathcal{L}_{1/2}(2k)$.

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

Assume that $L^+(2k) \neq \emptyset$. Then by definition of $L^+(2k)$ there exist $w \in L$, some $v \in A^*$ and $w_0, w_1, \ldots, w_{2k} \in A^*$ such that $w = w_0 \preceq w_1 \preceq w_2 \preceq \ldots \preceq w_{2k} \preceq v$ with $w_{2i} \in L$ and $w_{2i-1} \notin L$. For any $i \in \{0, 1, \ldots, 2k-1\}$ there must be two indices $j, j' \in \{0, \ldots, 2k\}$ 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 Theorem 2.3 that $\langle L_j \rangle_{\preceq} = L_j$ for all j. So from $w_i \preceq w_{i+1}$ we can conclude that $w_{i+1} \in L_j$ as well, which implies $j' > j$. Consequently, the words $w_0, w_1, \ldots, w_{2k}$ are in $2k+1$ 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 $2k$ such sets. Hence $m^+(L) < 2k$. ❑

In the remainder of this section we will give two applications of the above criterion for membership in $\mathcal{L}_{1/2}(k)$. First, we can conclude that the boolean hierarchy over $\mathcal{L}_{1/2}$ is a proper hierarchy.

Theorem 4.3. *For every $k \geq 1$,*

$$\mathcal{L}_{1/2}(k) \subsetneq \mathcal{L}_{1/2}(k+1).$$

Proof. Fix some $a \in A$, and define $|w|_a$ to be the number of occurences of a in $w \in A^*$. For $k \geq 1$ define

1. $M_{2k-1} =_{\mathrm{def}} \big\{ w \in A^* \;\big|\; |w|_a \text{ is odd or } |w|_a > 2k-1 \big\}$, and
2. $M_{2k} =_{\mathrm{def}} \big\{ w \in A^* \;\big|\; |w|_a \text{ is odd and } |w|_a \leq 2k \big\}$.

Obviously it holds that $m^-(M_k) = k$ and $m^+(M_k) = k-1$. By Theorem 4.2 we obtain $M_k \in \mathcal{L}_{1/2}(k) \backslash \mathrm{co}\mathcal{L}_{1/2}(k)$, and by Lemma 2.1.2 we get $\mathcal{L}_{1/2}(k) \subsetneq \mathcal{L}_{1/2}(k+1)$. ❑

Next we consider the decidability of the classes $\mathcal{L}_{1/2}(k)$. For a given dfa F, the equivalence $L(F) \in \mathcal{L}_{1/2}(k) \Leftrightarrow m^+(L(F)) < k$ given by Theorem 4.2 can be used to obtain a decision procedure for the question $L(F) \overset{?}{\in} \mathcal{L}_{1/2}(k)$. This follows from the next lemma. Here and in the sequel we assume that a regular language is given by a deterministic finite automaton.

Lemma 4.4. *Given a dfa F and $k \geq 1$, the questions $m^+(L(F)) \overset{?}{<} k$ and $m^-(L(F)) \overset{?}{<} k$ are decidable in nondeterministic space $k \cdot \log |F|$.*

Proof. Note that $m^+(L(F)) < k \Leftrightarrow L(F)^+(k) = \emptyset \Leftrightarrow L(F_k) = \emptyset$ where F_k is the nfa constructed in the proof of Lemma 3.5. Obviously, $L(F_k) = \emptyset$ is equivalent with the non-existence of a path between the starting state of F_k and one of its accepting states. Hence, we have to solve the graph non-accessibility problem for the transition graph of F_k which is of size $|A| \cdot |F|^{k+1}$. This can be done in co-nondeterminstic space $\log(|F|^{k+1}) = (k+1) \cdot \log |F|$ which is the same as nondeterministic space $k \cdot \log |F|$ [Imm88, Sze87]. ❑

Theorem 4.5. *For fixed $k \geq 1$, the decision problems for $\mathcal{L}_{1/2}(k)$ and $\mathrm{co}\mathcal{L}_{1/2}(k)$ are in* NL.

We are able to decide the question $m^+(L(F)) \overset{?}{<} k$ for given dfa F and $k \geq 1$. However, this does not mean automatically that we are able to compute $m^+(L(F))$ effectively. That this is indeed possible can be concluded from the following dichotomy-lemma by J. Stern.

Lemma 4.6 [Ste85]. *For a deterministic finite automaton F,*

$$m^+(L(F)) < \infty \Longleftrightarrow m^+(L(F)) \leq 2^{|F| \cdot |A|^2}.$$

This dichotomy enables us to compute the measure $m^+(L(F))$ simply by deciding the questions $m^+(L(F)) \overset{?}{<} k$ for $k = 1, 2, \ldots, 2^{|F| \cdot |A|^2} + 1$ with help of Lemma 4.4.

Theorem 4.7. *The measures $m^+(L)$ and $m^-(L)$ for a regular language L are computable in space $2^{\mathcal{O}(|F|)}$.*

Due to the close connection to the $\mathrm{FO}[<]$-logic (Theorem 2.2) we immediately have the following corollary.

Corollary 4.8. *The classes of the boolean hierarchy over the class Σ_1 of $\mathrm{FO}[<]$-logic are decidable.*

5 A Pattern Characterization for $\mathcal{L}_1$

In this section we give a "forbidden-pattern" characterization of the class $\mathcal{L}_1$ (for other characterizations of this class see [Sim75, Ste85]). First we define significant patterns that lead to infinite alternating extension chains. The technically involved part in the proof of the following theorem is to show conversely that an infinite alternating extension chain implies the occurence of such a pattern. For this end we continuously select suitable infinite subchains of an infinite chain, we emphasize on the position in a word where insertion of a letter leads to alternation and we extensively exploit the finiteness of an automaton.

We say that the dfa $F = (A, S, \delta, s_0, S')$ has the pattern P_1 (cf. Figure 1) if there exist $v, x, y, z \in A^*, a \in A$ and states $s_1, s_2, s_3 \in S$ such that $ya \preceq v$, $\delta(s_0, x) = \delta(s_1, v) = s_1$, $\delta(s_1, y) = s_2$, $\delta(s_2, a) = s_3$ and $\delta(s_2, z) \in S' \Leftrightarrow \delta(s_3, z) \notin S'$.

We say that the dfa F has the pattern P_2 (cf. Figure 2) if there exist $u, x, z, z' \in A^*, a \in A$ and states $s_1, s_2, s_3, s_4 \in S$ such that $az \preceq u$, $\delta(s_0, x) = s_1$, $\delta(s_1, a) = s_2$, $\delta(s_1, z) = \delta(s_3, u) = s_3$, $\delta(s_2, z) = \delta(s_4, u) = s_4$ and $\delta(s_3, z') \in S' \Leftrightarrow \delta(s_4, z') \notin S'$.

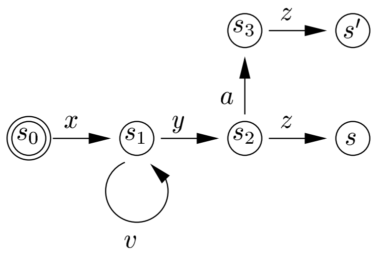

Figure 1: Pattern P_1 with $ya \preceq v$ and $s \in S' \Leftrightarrow s' \notin S'$.

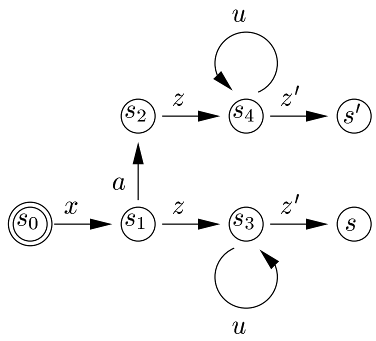

Figure 2: Pattern P_2 with $az \preceq u$ and $s \in S' \Leftrightarrow s' \notin S'$.

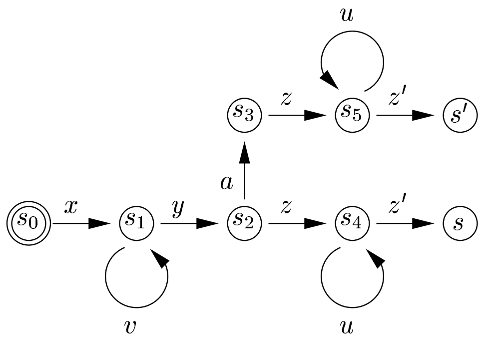

Figure 3: Pattern P_3 with $ya \preceq v$ or $az \preceq u$, and $s \in S' \Leftrightarrow s' \notin S'$.

We say that the dfa F has the pattern P_3 (cf. Figure 3) if there exist $u, v, x, y, z, z' \in A^*, a \in A$ and states $s_1, s_2, s_3, s_4, s_5 \in S$ such that $ya \preceq v$ or $az \preceq u$, $\delta(s_0, x) = \delta(s_1, v) = s_1$, $\delta(s_1, y) = s_2$, $\delta(s_2, a) = s_3$, $\delta(s_2, z) = \delta(s_4, u) = s_4$, $\delta(s_3, z) = \delta(s_5, u) = s_5$ and $\delta(s_4, z') \in S' \Leftrightarrow \delta(s_5, z') \notin S'$.

Theorem 5.1. *Let F be a dfa and let $\hat{F}$ be a dfa such that $L(F)^R = L(\hat{F})$. Then the following are equivalent:*

(1) $L(F) \in \mathcal{L}_1$,

(2) neither F nor $\hat{F}$ does have the pattern P_1,

(3) neither F nor $\hat{F}$ does have the pattern P_2,

(4) F does not have the pattern P_3.

In the proof we will make use of the following easy to see lemma.

Lemma 5.2. *Let $\{\alpha_i\}$ be a sequence of real numbers such that $0 < \alpha_i < 1$ and $\alpha_i \neq \alpha_j$ for $i \neq j$. Then there exists an infinite monotonic subsequence of $\{\alpha_i\}$.*

Proof of Theorem 5.1. (2) $\Rightarrow$ (1): Assume that $L(F) \notin \mathcal{L}_1$ for some dfa $F = (A, S, \delta, s_0, S')$. We have to show that F has pattern P_1 or any $\hat{F}$ with $L(F)^R = L(\hat{F})$ has pattern P_2. First we conclude with Theorem 4.1 that $m^+(L(F))$ is infinite and we can assume w.l.o.g. that there exists an infinite sequence of words $\{w_j\}$ and a letter $a \in A$ such that $w_j \preceq w_{j+1}$ for all $j \geq 1$, and $w_{2i-1} = w_i'w_i''$, $w_{2i} = w_i'aw_i''$, $\delta(s_0, w_{2i-1}) \notin S'$ and $\delta(s_0, w_{2i}) \in S'$ for all $i \geq 1$. Next we introduce markers m_i at the positions where a is inserted when going from w_{2i-1} to w_{2i}, i.e. the word $w_i'aw_i''$ has markers $m_1, m_2, \ldots, m_i$. To show the existence of an infinite subsequence of words which is monotonic with respect to the insertion positions of the letter a, we inductively attach values $\alpha_i \in \mathbb{R}$ to each marker m_i as follows: Let $\alpha_{i+1} =_{\mathrm{def}} (\beta_{i+1} + \gamma_{i+1})/2$ with $\beta_{i+1} =_{\mathrm{def}} \max\left(\left\{ \alpha_j \;\middle|\; 1 \leq j \leq i \text{ and marker } m_j \text{ is left to } m_{i+1} \right\} \cup \{0\}\right)$ and $\gamma_{i+1} =_{\mathrm{def}} \min\left(\left\{ \alpha_j \;\middle|\; 1 \leq j \leq i \text{ and marker } m_j \text{ is right to } m_{i+1} \right\} \cup \{1\}\right)$. We observe that m_i is left to m_j if and only if $\alpha_i < \alpha_j$. Now Lemma 5.2 tells us that there is an infinite strictly monotonic subsequence of $\{\alpha_i\}$. We distinguish two cases.

Case 1. Assume that there exists an infinite strictly increasing subsequence of $\{\alpha_i\}$, i.e. there is a mapping $\tau : \mathbb{N} \to \mathbb{N}$ such that $\tau(i) < \tau(i+1)$ and $\alpha_{\tau(i)} < \alpha_{\tau(i+1)}$ for all $i \geq 1$. For simplicity we redefine $w_i =_{\mathrm{def}} w_{\tau(i)}$ and summarize the properties of the sequence selected in this way. For all $i \geq 1$ we have

1. $w_{2i-1} = w_i'w_i'' \preceq w_i'aw_i'' = w_{2i} \preceq w_{2i+1} = w_{i+1}'w_{i+1}''$,
2. $w_i'a \preceq w_{i+1}'$,
3. $\delta(s_0, w_{2i-1}) \notin S'$ and $\delta(s_0, w_{2i}) \in S'$.

We use the sequence $\{w_i'\}$ as a starting point for subsequent selections of sequences $\{w_{i,k}\}$ for $k = 0, 1, 2, \ldots$ all having the properties stated in the following claim. Using the finiteness of the set of states will then enable us to find the pattern P_1 in F. In the following notations a superscript in combination with a subscript denotes an index.

Claim. For every $k \geq 0$ there exists a state $s_k \in S$ and an infinite subsequence $\{w_{i,k}\}$ of $\{w_i'\}$ such that for all k, i there are words $v_{i,k}^1, v_{i,k}^2, \ldots, v_{i,k}^k, u_{i,k} \in A^*$ with

a. $w_{i,k} = v_{i,k}^1 a v_{i,k}^2 a \cdots a v_{i,k}^k a u_{i,k}$,

b. $v_{i,k}^j \preceq v_{i+1,k}^j$ for $1 \leq j \leq k$,

c. $u_{1,k-1} \preceq v_{i,k}^k$ for $k \geq 1$,

d. $u_{i,k}a \preceq u_{i+1,k}$, and

e. $\delta(s_0, v_{i,k}^1 a v_{i,k}^2 a \cdots a v_{i,k}^j a) = s_j$ for $1 \leq j \leq k$.

Proof of claim. We proceed by induction on k. The case $k = 0$ is easy to see with $w_{i,0} =_{\text{def}} u_{i,0} = w_i^{'}$. Starting with $\{w_{i,k}\}$ we show how to select a subsequence $\{w_{i,k+1}\}$ fulfilling the assertions of the claim. First we observe that we can conclude from $u_{i,k}a \preceq u_{i+1,k}$ that $u_{1,k}a \preceq u_{i,k}$ for all $i \geq 2$. Now for every $i \geq 2$ we can identify in $u_{i,k}$ a word left (right, resp.) of this particular letter a, i.e. there are words $v_{i,k}^{k+1}$ and $u_{i,k}^{'}$ such that $u_{i,k} = v_{i,k}^{k+1}au_{i,k}^{'}$, $u_{1,k} \preceq v_{i,k}^{k+1} \preceq v_{i+1,k}^{k+1}$ and $u_{i,k}^{'}a \preceq u_{i+1,k}^{'}$. Hence we can write each $w_{i,k}$ as $w_{i,k} = v_{i,k}^1 av_{i,k}^2 a \cdots av_{i,k}^k av_{i,k}^{k+1} au_{i,k}^{'}$. Due to the finiteness of the set of states of F we can conclude that there exists a state $s_{k+1} \in S$ and a strictly increasing mapping $\tau : \mathbb{N} \to \mathbb{N}$ such that $\delta(s_0, v_{\tau(i),k}^1 av_{\tau(i),k}^2 a \cdots av_{\tau(i),k}^k av_{\tau(i),k}^{k+1} a) = s_{k+1}$. Now we define $w_{i,k+1} =_{\text{def}} w_{\tau(i),k}$, $v_{i,k+1}^j =_{\text{def}} v_{\tau(i),k}^j$ for $1 \leq j \leq k+1$ and $u_{i,k+1} =_{\text{def}} u_{\tau(i),k}^{'}$. We leave the verification of the assertions a to e for $\{w_{i,k+1}\}$ as an exercise. *(End proof of claim)*

We keep the notations of the claim. Now, again due to the finiteness of S there exist k, m with $1 \leq k < m \leq |S| + 1$ and $s_k = s_m$. Hence we can define $x =_{\text{def}} v_{1,m-1}^1 av_{1,m-1}^2 a \cdots av_{1,m-1}^k a$, $v =_{\text{def}} v_{1,m}^{k+1} av_{1,m}^{k+2} a \cdots av_{1,m}^m a$, $y =_{\text{def}} v_{1,m-1}^{k+1} av_{1,m-1}^{k+2} a \cdots av_{1,m-1}^{m-1} au_{1,m-1}$, and $z =_{\text{def}} w_r^{''}$, where r is the index such that $w_{1,m-1} = w_r^{'}$. Note that $xy = w_r^{'}$. We conclude with the assertions of the claim, that

$$\begin{aligned} ya &= v_{1,m-1}^{k+1} a \cdots av_{1,m-1}^{m-1} au_{1,m-1} a \\ &\preceq v_{\tau(1),m-1}^{k+1} a \cdots av_{\tau(1),m-1}^{m-1} av_{1,m}^m a \\ &= v_{1,m}^{k+1} a \cdots av_{1,m}^{m-1} av_{1,m}^m a \\ &= v \end{aligned}$$

Moreover we see that $\delta(s_0, xyz) = \delta(s_0, w_r^{'}w_r^{''}) = \delta(s_0, w_{2r-1}) \notin S^{'}$ and $\delta(s_0, xyaz) = \delta(s_0, w_r^{'}aw_r^{''}) = \delta(s_0, w_{2r}) \in S^{'}$. This shows that F has pattern P_1.

Case 2. Now assume that there exists an infinite strictly decreasing subsequence of $\{\alpha_i\}$. Then obviously $\{w_j^R\}$ is an infinite alternating extension chain with respect to $L(F)^R$. Let $\hat{F}$ be a dfa accepting $L(F)^R$. Attaching markers $\alpha_i^{'}$ in the same way as above leads to $\alpha_i^{'} = 1 - \alpha_i$ and hence there is a strictly increasing subsequence of $\{\alpha_i^{'}\}$. We can conclude as in case 1 that $\hat{F}$ has pattern P_1. This finishes the proof of $(2) \Rightarrow (1)$ and we turn to the remaining implications.

$(1) \Rightarrow (4)$: Suppose some dfa F has pattern P_3. Then we have for $i \geq 0$ the infinite alternating word extension chain $xv^iyzu^iz^{'} \preceq xv^iyazu^iz^{'} \preceq xv^{i+1}yzu^{i+1}z^{'}$ since either $ya \preceq v$ or $az \preceq u$.

$(4) \Rightarrow (3)$: If some dfa F has pattern P_2 then this is also a pattern P_3 (with $v = y = \epsilon$), which is a contradiction. Next we show that if some dfa $\hat{F}$ has pattern P_2 then any dfa F with $L(F) = L(\hat{F})^R$ has pattern P_1, and again this is also pattern P_3 (with $u = z^{'} = \epsilon$), a contradiction as well. So suppose that a dfa $\hat{F} = (A, \hat{S}, \hat{\delta}, \hat{s}_0, \hat{S}^{'})$ has the pattern P_2 witnessed by $x, z, u, z^{'} \in A^*$ and $a \in A$. Let $F = (A, S, \delta, s_0, S^{'})$ be any dfa with $L(F) = L(\hat{F})^R$ and choose $m, k \in \mathbb{N}$ with $m > k \geq 0$ such that $\delta\big(s_0, (z^{'})^R(u^R)^k\big) = \delta\big(s_0, (z^{'})^R(u^R)^{k+m}\big)$. We define $\bar{x} =_{\text{def}} (z^{'})^R(u^R)^k$, $\bar{v} =_{\text{def}} (u^R)^m$, $\bar{y} =_{\text{def}} z^R$ and $\bar{z} =_{\text{def}} x^R$. Now one can easily verify that $\bar{x}, \bar{v}, \bar{y}, \bar{z} \in A^*$ and $a \in A$ give rise to pattern P_1 in F since $\bar{y}a \preceq \bar{v}$ follows from $az \preceq u$.

$(3) \Rightarrow (2)$. Suppose that a dfa $F = (A, S, \delta, s_0, S^{'})$ has pattern P_1 witnessed by $x, v, y, z \in A^*$ and $a \in A$. Let $\hat{F} = (A, \hat{S}, \hat{\delta}, \hat{s}_0, \hat{S}^{'})$ be any dfa with $L(\hat{F}) = L(F)^R$ and choose $m, k \in \mathbb{N}$ with $m > k \geq 0$ such that $\hat{\delta}\big(\hat{s}_0, (yz)^R(v^R)^k\big) = \hat{\delta}\big(\hat{s}_0, (yz)^R(v^R)^{k+m}\big)$ and $\hat{\delta}\big(\hat{s}_0, (yaz)^R(v^R)^k\big) = \hat{\delta}\big(\hat{s}_0, (yaz)^R(v^R)^{k+m}\big)$. We define $\bar{x} =_{\text{def}} z^R$, $\bar{u} =_{\text{def}} (v^R)^m$, $\bar{z} =_{\text{def}} y^R(v^R)^k$ and $\bar{z}^{'} =_{\text{def}} x^R$.

Again, one can easily verify that $\bar{x}, \bar{u}, \bar{z}, \bar{z}' \in A^*$ and $a \in A$ give rise to pattern P_2 in $\hat{F}$ since $a\bar{z} \preceq \bar{u}$ follows from $ya \preceq v$. ❑

We remark that the proof of $(2) \Rightarrow (1)$ even shows that the automata F and $\hat{F}$ do not have the *two* instances of pattern P_1 with $s \in S'$ and $s' \notin S'$ on one hand, and $s' \in S'$ and $s \notin S'$ on the other hand. The same holds analogously for the other patterns. To see this note that we can start the whole investigation at the very beginning of the proof with the sequence $\{w_{j+1}\}$.

Using the above Theorem we obtain a co-NL(=NL)-algorithm for the decision problem for $\mathcal{L}_1$ simply by testing the occurence of the pattern P_3 in a given dfa. This algorithm is completely different from those which follow from the characterizations in [Sim75, Ste85]. Note that S. Cho and D.T. Huynh proved in [CH91] that the decision problem for $\mathcal{L}_1$ is even NL-complete.

6 Complexity Theoretical Consequences

Let a nondeterministic polynomial time Turing machine M output on every path a symbol from A and assume a fixed ordering on the set of all paths. We additionally assume here that, given some input x and the number of a path i, one can compute in polyomial time the output of M on path i (*balanced* computation tree). This leads in a natural way to the notion of the *leafstring* of M on some input x when concatenating the output symbols of M's computation tree. Now a language $L \subseteq A^*$ gives rise to the class $\mathrm{Leaf}^{\mathrm{P}}(L)$ of all languages L' for which there is a machine M of the above type such that for all x it holds that $x \in L'$ if and only if the leafstring of M on input x belongs to L. Furthermore, for some class $\mathcal{C}$, denote by $\mathrm{Leaf}^{\mathrm{P}}(\mathcal{C})$ the union of all classes $\mathrm{Leaf}^{\mathrm{P}}(L)$ with $L \in \mathcal{C}$.

As stated in the introduction this leaf language approach led to new insights into the structure of complexity classes between P and PSPACE. However, most results deal with *classes* of leaf languages and an important question is what complexity classes are definable by a *single* leaf language. Some progress in this direction has been made in [Bor95, BKS98].

Due to the close connection of the classes of the Straubing-Thérien hierarchy to FO[<]-logic (Theorem 2.2) we can make use of the known relationship between languages definable within this logic and the classes of the polynomial time hierarchy.

Theorem 6.1 [BV98]. *Let A be an arbitrary alphabet with $|A| \geq 2$ and let $k \geq 1$.*

1. $\Sigma_k^{\mathrm{p}} = \mathrm{Leaf}^{\mathrm{P}}(\mathcal{L}_{k-1/2})$
2. $\Pi_k^{\mathrm{p}} = \mathrm{Leaf}^{\mathrm{P}}(\mathrm{co}\mathcal{L}_{k-1/2})$

The "forbidden-pattern" characterization of the classes $\mathcal{L}_{1/2}$ from [PW97] enables us to show which complexity classes are exactly definable by a single leaf language from this class.

Theorem 6.2. *For an arbitrary alphabet A with $|A| \geq 2$ we have*

$$\left\{ \mathrm{Leaf}^{\mathrm{P}}(L) \;\middle|\; L \in \mathcal{L}_{1/2} \right\} = \left\{ \{\emptyset\}, \{B^* \mid B \text{ finite alphabet}\}, \mathrm{P}, \mathrm{NP} \right\}$$

and given some dfa accepting a language $L \in \mathcal{L}_{1/2}$ one can effectively determine the class on the right hand side with which $\mathrm{Leaf}^{\mathrm{P}}(L)$ coincides.

For single leaf languages from the boolean hierarchy over $\mathcal{L}_{1/2}$ the situation is a lot more complicated. However, we have the following "union-style" theorem which provides an upper bound for complexity classes definable via such leaf languages. Throughout the paper we studied the classes $\mathcal{L}_{1/2}(k)$ for an arbitrary but fixed alphabet A. Now we will emphasize on the chosen alphabet and denote by $\mathcal{L}^A_{1/2}(k)$ the classes $\mathcal{L}_{1/2}(k)$ defined for languages over A.

Theorem 6.3. *For any $k \geq 1$,*

$$\mathrm{NP}(k) \quad = \bigcup_{A \text{ finite alphabet}} \mathrm{Leaf}^{\mathrm{P}}\big(\mathcal{L}^A_{1/2}(k)\big).$$

Proof. To see the inclusion from right to left note with Theorem 6.1.1 that $\mathrm{Leaf}^{\mathrm{P}}(\mathcal{L}^A_{1/2}) \subseteq \mathrm{NP}$ for any alphabet A. Furthermore it holds for languages L_1, L_2 that $\mathrm{Leaf}^{\mathrm{P}}(L_1 \cup L_2) \subseteq \mathrm{Leaf}^{\mathrm{P}}(L_1) \vee \mathrm{Leaf}^{\mathrm{P}}(L_2)$, $\mathrm{Leaf}^{\mathrm{P}}(L_1 \cap L_2) \subseteq \mathrm{Leaf}^{\mathrm{P}}(L_1) \wedge \mathrm{Leaf}^{\mathrm{P}}(L_2)$ and $\mathrm{Leaf}^{\mathrm{P}}(\overline{L_1}) = \mathrm{coLeaf}^{\mathrm{P}}(L_1)$, where $\mathcal{C}_1 \vee \mathcal{C}_2 =_{\mathrm{def}} \big\{\, L'_1 \cup L'_2 \;\big|\; L'_1 \in \mathcal{C}_1, L'_2 \in \mathcal{C}_2 \,\big\}$ and $\mathcal{C}_1 \wedge \mathcal{C}_2 =_{\mathrm{def}} \big\{\, L'_1 \cap L'_2 \;\big|\; L'_1 \in \mathcal{C}_1, L'_2 \in \mathcal{C}_2 \,\big\}$ for classes $\mathcal{C}_1, \mathcal{C}_2$.

For the other inclusion define for $k \geq 1$ the alphabet $A_k =_{\mathrm{def}} \{0, 1, 2, \ldots, k\}$ and the language $L_k =_{\mathrm{def}} \big\{\, w \in A_k^* \;\big|\; \max\{i \in A_k \mid i \preceq w\} \text{ is odd} \,\big\}$. One can show with Lemma 2.1 that $\mathrm{Leaf}^{\mathrm{P}}(L_k) = \mathrm{NP}(k)$. Observe that $m^+(L_k) = k - 1$, so with Theorem 4.2 it follows that $L_k \in \mathcal{L}^{A_k}_{1/2}(k)$. ❑

Corollary 6.4. *If $m^+(L) < k$ for a regular language L then* $\mathrm{Leaf}^{\mathrm{P}}(L) \subseteq \mathrm{NP}(k)$.

Note that the measure m^+ is computable (Theorem 4.7). Moreover the results obtained here remain valid if we omit the restriction that the computation tree of a Turing machine must be balanced.

Finally we compare our results with related work. In [CHVW98] the case of commutative leaf languages has been studied, i.e. the case where membership to a language depends only on the numbers of occurences of the alphabet symbols. For an oracle D we denote by $\mathcal{C}^D$ the relativized version of a complexity class $\mathcal{C}$. It has been proved in the mentioned paper that for every commutative language L,

$$m^+(L) < k \iff \forall D\big(\mathrm{Leaf}^{\mathrm{P}}(L)^D \subseteq \mathrm{NP}(k)^D\big).$$

Furthermore, other (stronger) measures n^+ and n^- have been defined, i.e. $n^+(L) \leq m^+(L)$ and $n^-(L) \leq m^-(L)$, and it has been proved that for every commutative language L,

$$n^-(L) \geq k \iff \forall D\big(\mathrm{Leaf}^{\mathrm{P}}(L)^D \supseteq \mathrm{NP}(k)^D\big).$$

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.

[BCS92] D. P. Bovet, P. Crescenzi, and R. Silvestri. A uniform approach to define complexity classes. *Theoretical Computer Science*, 104:263–283, 1992.

[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 *Proceedings 25rd International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science. Springer Verlag, 1998. To appear.

[Bor95] B. Borchert. On the acceptance power of regular languages. *Theoretical Computer Science*, 148:207–225, 1995.

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

[BV98] H.-J. Burtschick and H. Vollmer. Lindström quantifiers and leaf language definability. *International Journal of Foundations of Computer Science*, 1998. To appear.

[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.

[CHVW98] K. Cronauer, U. Hertrampf, H. Vollmer, and K. W. Wagner. The chain method to separate counting classes. *Theoretical Computer Science*, 31:93–108, 1998.

[HLS$^+$93] U. Hertrampf, C. Lautemann, T. Schwentick, H. Vollmer, and K. W. Wagner. On the power of polynomial time bit-reductions. In *Proceedings 8th Structure in Complexity Theory*, pages 200–207, 1993.

[Imm88] N. Immerman. Nondeterministic space is closed under complementation. *SIAM Journal on Computing*, 17:935–938, 1988.

[JMT94] B. Jenner, P. McKenzie, and D. Thérien. Logspace and logtime leaf languages. In *9th Annual Conference Structure in Complexity Theory*, pages 242–254, 1994.

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

[MP71] R. McNaughton and S. Papert. *Counterfree Automata*. MIT Press, Cambridge, 1971.

[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.

[PP86] D. Perrin and J. E. Pin. First-order logic and star-free sets. *Journal of Computer and System Sciences*, 32:393–406, 1986.

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

[Sim75] I. Simon. Piecewise testable events. In *Proceedings 2nd GI Conference*, volume 33 of *Lecture Notes in Computer Science*, pages 214–222. Springer-Verlag, 1975.

[Ste85] J. Stern. Characterizations of some classes of regular events. *Theoretical Computer Science*, 35:17–42, 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.

[Sze87] R. Szelepcsényi. The method of forcing for nondeterministic automata. *Bull.of the EATCS*, 33:96–100, 1987.

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

[Tho82] W. Thomas. Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25:360–376, 1982.

[Tho96] W. Thomas. Languages, automata, and logic. Technical Report 9607, Institut für Informatik und praktische Mathematik, Universität Kiel, 1996.

[Ver93] N. K. Vereshchagin. Relativizable and non-relativizable theorems in the polynomial theory of algorithms. *Izvestija Rossijskoj Akademii Nauk*, 57:51–90, 1993. In Russian.