OPERATOR PRECEDENCE LANGUAGES: THEIR AUTOMATA-THEORETIC AND LOGIC CHARACTERIZATION∗

VIOLETTA LONATI†, DINO MANDRIOLI‡, FEDERICA PANELLA‡, AND MATTEO PRADELLA§

Abstract. Operator precedence languages were introduced half a century ago by Robert Floyd to support deterministic and efficient parsing of context-free languages. Recently, we renewed our interest in this class of languages thanks to a few distinguishing properties that make them attractive for exploiting various modern technologies. Precisely, their local parsability enables parallel and incremental parsing, whereas their closure properties make them amenable to automatic verification techniques, including model checking. In this paper we provide a fairly complete theory of this class of languages: we introduce a class of automata with the same recognizing power as the generative power of their grammars; we provide a characterization of their sentences in terms of monadic second-order logic as has been done in previous literature for more restricted language classes such as regular, parenthesis, and input-driven ones; we investigate preserved and lost properties when extending the language sentences from finite length to infinite length (ω-languages). As a result, we obtain a class of languages that enjoys many of the nice properties of regular languages (closure and decidability properties, logic characterization) but is considerably larger than other families—typically parenthesis and input-driven ones—with the same properties, covering “almost” all deterministic languages.

Key words. operator precedence, visibly pushdown languages, monadic second-order logic, ω-languages

AMS subject classifications. 03D05, 68Q45

DOI. 10.1137/140978818

Introduction. Operator precedence grammars (OPGs) and languages (OPLs) certainly deserve an important place in the history of formal languages and compilers. They were invented by Floyd [23] with the major motivation of enabling efficient, deterministic parsing of programming languages. In fact Floyd’s intuition was inspired by arithmetic expressions whose structure is determined either by explicit parentheses or by the conventional, “hidden” precedence of multiplicative operators over additive ones. By generalizing this observation Floyd defined three basic relations between terminal symbols, namely, yields and takes precedence and equal in precedence (resp., denoted by symbols $\preceq$, $\succeq$, $\equiv$), in such a way that the right-hand side (r.h.s.) of an OPG rule is enclosed within a pair $\preceq$, $\succeq$, and $\equiv$ holds between consecutive terminal symbols thereof (in OPGs nonterminal symbols are “transparent,” i.e., irrelevant, w.r.t. the precedence relations [23]).

Subsequently, under the main motivation of grammar inference, it was shown that, once an operator precedence matrix (OPM) is given such that at most one relation holds between any two terminal characters, the family of OPLs sharing the given OPM is a Boolean algebra [19]. This result somewhat generalizes closure properties enjoyed by regular languages and by context-free (CF) languages whose structure, i.e.,
the syntax tree, is immediately visible in the terminal sentences, such as parenthesis languages [31] and tree-automata languages [11]. Such interesting algebraic properties enabled original inference algorithms, such as those proposed in [20]. After these initial results the theoretical investigation of OPLs was almost abandoned, most likely because of the advent of more general grammars, mainly the left-to-right rightmost (LR) ones [26], which support parsing algorithms for the whole class of deterministic CF languages. Nevertheless OPG-based parsing remains of some interest thanks to its simplicity and efficiency and is still used, at least partially, in many practical cases [24].

In the last decades, instead, an independent branch of research generated a flourishing of new results in terms of logic characterization of language families, ignited by the pioneering results by Büchi [12] and Muller [32] on the monadic second-order (MSO) logic characterization of regular languages over finite or infinite words (ω-languages) and motivated mainly by the breakthrough application of model checking, which is rooted in closure properties and decidability of the emptiness problem, besides correspondence between automata-theoretic and logic language characterization. The present state of the art exhibits plenty of language families and related characterization in terms of various forms of logic formalisms (first order, propositional, temporal logic, and more specialized ones [17, 11]); most of them are motivated by the wish to extend model-checking techniques, i.e., decidability of system properties, beyond the natural scope of finite-state machines.

Within such a rich literature, visibly pushdown languages (VPLs) [4], previously known as input-driven languages (IDLs) [5], certainly deserve a major role. In a nutshell IDLs alias VPLs are based on, and extend, original parenthesis languages [31], e.g., by allowing for unmatched closed and open parentheses at the beginning and end of a sentence, respectively. Throughout the years this research field produced a fairly complete study of this family of languages whose main features can be summarized as follows:

- Being essentially a generalization of parenthesis languages their structure is immediately transparent at the “surface sentence,” unlike more general CF languages; arithmetic expressions, e.g., which are found in practically every programming language, do not reflect in the sequence of the leaves of the syntax tree the internal structure of the tree, which can be built only by knowing that multiplication operators take precedence over the additive ones.
- They have a complete characterization in terms of pushdown-automata and CF grammar families recognizing and generating, them, respectively.
- They are closed w.r.t. to all fundamental language operations (Boolean, concatenation, Kleene *, . . .), like regular languages and unlike more general CF families.
- Within the landscape of algorithms that are necessary to develop model-checking techniques—whose complexities span from NP to PSPACE, EXPTIME, . . . completeness—they exhibit “comparable” complexities: for instance, the core algorithm for determinizing nondeterministic visibly pushdown automata (VPAs) has $2^{O(s^2)}$ complexity w.r.t. the cardinality $s$ of the original state space and the inclusion problem for VPLs of both finite and infinite strings is EXPTIME-complete.
- They are characterized in terms of a MSO logic that applies both to finite and infinite-length words.

Similar results have been obtained also for other classes of languages on the basis of the strong motivation provided by “model-checking-like” applications [10].
Recently, our interest in OPLs has been renewed thanks to two, seemingly unrelated, properties thereof. The first one is their local parsability, i.e., the fact that the typical shift-reduce parsing algorithm associated with them, determines the replacing of a r.h.s. by the corresponding left-hand side (l.h.s.) exclusively on the basis of the embracing $<$ and $>$ relations, i.e., independently on parts of the string that may be arbitrarily far from the considered r.h.s. This property is not enjoyed by more powerful grammars such as LR ones and nowadays it may compensate for the minor loss of generative power because it makes it easier and more efficient to exploit parallelism and incrementality in the parsing of large strings formalizing complex systems and their behavior. The exploitation of this property, however, is the target of a different and—so far— independent research whose first results are documented in [7] and [6].

In this paper, instead, we focus on another, equally stimulating property of OPLs. We realized, in fact, that the OPL family strictly includes the independently studied family of VPLs and other related ones such as balanced languages [8]. On the basis of this somewhat surprising remark we further investigated other closure properties of OPLs besides the Boolean ones that were originally proved in [19]: the result is that OPLs are, to the best of our knowledge, the largest class of languages that enjoys all major closure properties that are typical of regular languages [18].\(^1\)

Herewith the goal of this paper: in order to apply to OPLs the same successful verification techniques formerly developed for regular languages, VPLs, and other—input-driven—language families, we develop a complete automata-theoretic and logic characterization of OPLs. In fact resuming the study of this old family of languages showed unexpected similarities with, and generalizations of, the peculiar properties of seemingly unrelated and differently motivated classes of languages.

In our opinion OPLs offer a surprising combination of the merits of IDLs and of those of more general deterministic CF languages. On the one hand, in fact, they are input driven since their analysis can be based exclusively on the input characters and their pairwise relations; but, unlike more traditional IDLs, they are well suited to formalize general programming languages and other languages of practical interest; such a distinguishing feature allows us to extend to them closure and decidability properties not enjoyed by more general CF languages. On the other hand, their minor lack of power w.r.t. deterministic languages does not prevent them from including most programming languages of practical interest: previous efforts in fact produced compilers based on OPGs for various programming languages such as ALGOL 68 and Prolog [23, 21]; more recently we exploited the mentioned property of local parsability to produce parallel parsers for JSON and Lua [6].

Given the fairly numerous collection of strongly connected properties, we structure the present paper into two main parts. The first one completes the path begun with [19] and resumed with [18] by providing a fairly complete theory of traditional OPLs defined on strings of finite length; precisely, we present the following:

- A new family of pushdown automata fully equivalent to OPGs; rather surprisingly, in fact, a precise automata-theoretic characterization of OPLs was missing in the original literature.\(^2\)
A complete characterization of OPLs in terms of MSO logic so as to align this family with a now classic approach of the literature—rooted in the work by Büchi. This allows, at least potentially, for the definition of model-checking algorithms to prove properties of languages defined either by means of generating grammars or by means of recognizing automata. Given the prohibitive complexity of decision algorithms based on MSO logic, however, it is common practice in the literature to resort to model-checking algorithms based on less powerful but simpler logics. We will provide a few hints on pursuing such an approach in the conclusion.

In the second part of this paper we define \(\omega\)OPLs, i.e., the OPLs of infinite words. Infinite word languages are becoming more and more relevant in the literature due to the need of modeling systems whose behavior proceeds indefinitely, such as operating systems, control systems, etc. After introducing and comparing various forms of acceptance of infinite words by our OPAs by paralleling classical literature of \(\omega\)-regular languages, we reinvestigate their main properties by pointing out which of them are preserved from the finite-length case and which ones are lost. This includes also a further characterization of \(\omega\)OPLs in terms of MSO logic.

In the conclusions we briefly hint at further research directions, noticeably investigating the relations of OPLs with less powerful but less complex logics than MSO ones, as it has been or is being done for other (input-driven) language families.

I: Finite Word Operator Precedence Languages.

This part is devoted to finite-length OPLs. After stating basic definitions and terminology (section 1) and resuming previous results already available in the open literature (section 1.1), we introduce the new class of pushdown automata explicitly tailored at OPLs: in section 2 we give the basic definitions and provide examples to show their usefulness in modeling various cases of practical interest; then we show the equivalence between deterministic and nondeterministic versions of these automata, at the price, however, of an increase in state space size given by an exponential function with quadratic exponent; we also study the complexity of decision problems for OPLs. Section 3 shows, in a constructive way, the equivalence between OPGs and the new class of automata; finally, section 4 presents an MSO logic characterization of OPLs.

1. Preliminaries. A CF grammar is a 4-tuple \(G = (N, \Sigma, P, S)\), where \(N\) is the nonterminal alphabet, \(\Sigma\) is the terminal one, \(P\) the rule (or production) set, and \(S \subseteq N\) the set of axioms.\(^3\) The empty string is denoted \(\varepsilon\). An empty rule has \(\varepsilon\) as the r.h.s. A renaming rule has one nonterminal as the r.h.s. A grammar is reduced if every rule can be used to generate some string in \(\Sigma^*\). It is invertible if no two rules have identical r.h.s’s.

The direct derivation relation is denoted by \(\Rightarrow\) and its reflexive transitive closure, the derivation relation, is denoted by \(\Rightarrow^*\).

The following naming convention will be adopted, unless otherwise specified: lowercase Latin letters \(a, b, \ldots\) denote terminal characters; uppercase Latin letters \(A, B, \ldots\) denote nonterminal characters; letters \(u, v, \ldots\) denote terminal strings; and Greek letters \(\alpha, \beta, \ldots\) denote strings over \(\Sigma \cup N\). The strings may be empty, unless stated otherwise.

\(^3\)This less usual but equivalent definition of axioms as a set has been adopted for parenthesis languages [31] and other IDLs; we chose it for this paper to simplify some notations and constructions.
In this initial part we will use arithmetic expressions, which are a small fraction of practically all programming languages, as a running example to introduce and explain the basic definitions, properties, and constructions referring to OPLs.

Example 1. Arithmetic expressions considered in this paper include two operators, an additive one and a multiplicative one that takes precedence over the other one, in the sense that, during the interpretation of the expression, multiplications must be executed before sums; as usual, parentheses are used to specify a different precedence hierarchy between the two operations. They are denoted by the special symbols \( [\text{ ]} \) to avoid overloading with the use of the same symbol in all other formulas of this paper. Figure 1 presents a grammar and the derivation tree of expression \( n + n \times (n + n) \) generated thereby; all nonterminals are also axioms.

Notice that the structure of the syntax tree (uniquely) corresponding to the input expression reflects the precedence order which drives computing the value attributed to the expression. This structure, however, is not immediately visible in the expression: in fact, Figure 2 proposes a different grammar which generates the same expressions as the grammar of Figure 1 but would associate with the same sentence the syntax tree displayed in the right part of the figure. Yet another (ambiguous) grammar could generate both. If instead we used a parenthesis grammar to generate
arithmetic expressions, it would produce the string \( \langle n + \langle n \times \langle n + n \rangle \rangle \rangle \) instead of the previous one and the structure of the corresponding tree would be immediately visible in the expression. For this reason we say that such general grammars “hide” the structure associated with a sentence—even when they are unambiguous—whereas parenthesis grammars and other input-driven ones make the structure explicit in the sentences they generate.

A rule is in operator form if its r.h.s. has no adjacent nonterminals; an operator grammar (OG) contains just such rules. Notice that, unlike the usual arithmetic relations denoted by similar symbols, the precedence relations do not enjoy any one of the transitive, symmetric, reflexive properties. For an OG above precedence relations do not enjoy any one of the transitive, symmetric, reflexive properties. For an OG

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>×</th>
<th>(</th>
<th>)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&lt;</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>×</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&lt;</td>
<td>&gt;</td>
<td>&lt;</td>
</tr>
<tr>
<td>(</td>
<td>&lt;</td>
<td>&lt;</td>
<td>≤</td>
<td>&lt;</td>
<td>≥</td>
</tr>
<tr>
<td>)</td>
<td>&gt;</td>
<td>&gt;</td>
<td>&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 displays the OPM associated with the grammar of Figure 1 where, for an OPG [23] a ref r o m[19] a n d[18], where they are also called Floyd grammars (FGs).

For an OG \( G \) and a nonterminal \( A \), the left and right terminal sets are

- left \( \mathcal{L}(A) = \{ a \in \Sigma \mid A \Rightarrow Baa \} \),
- right \( \mathcal{R}(A) = \{ a \in \Sigma \mid A \Rightarrow aaB \} \),

where \( B \in N \cup \{ \varepsilon \} \). The grammar name \( G \) will be omitted unless necessary to prevent confusion. For the grammar of Figure 1 the left and right terminal sets of nonterminals \( E, T \), and \( F \) are, respectively,

- \( \mathcal{L}(E) = \{ +, \times, n, () \} \),
- \( \mathcal{R}(E) = \{ +, \times, n, () \} \),
- \( \mathcal{L}(T) = \{ x, n, () \} \),
- \( \mathcal{R}(T) = \{ x, n, () \} \),
- \( \mathcal{L}(F) = \{ n, () \} \),
- \( \mathcal{R}(F) = \{ n, () \} \).

For an OG \( G \), let \( \alpha, \beta \) range over \( (N \cup \Sigma)^* \) and \( a, b \in \Sigma \). Three binary OP relations are defined:

- equal in precedence: \( a = b \iff \exists A : a \alpha aBb \beta, B \in N \cup \{ \varepsilon \} \),
- takes precedence: \( a \rightarrow b \iff \exists A : a \alpha bB, D \in N \) and \( a \in \mathcal{R}(D) \),
- yields precedence: \( a \leftarrow b \iff \exists A : a \alpha D \beta, D \in N \) and \( b \in \mathcal{L}(D) \).

Notice that, unlike the usual arithmetic relations denoted by similar symbols, the above precedence relations do not enjoy any one of the transitive, symmetric, reflexive properties. For an OG \( G \), the OPM \( M = OPM(G) \) is a \( |\Sigma| \times |\Sigma| \) array that, for each ordered pair \( (a, b) \), stores the set \( M_{ab} \) of OP relations holding between \( a \) and \( b \).

Table 1 displays the OPM associated with the grammar of Figure 1 where, for an ordered pair \( (a, b) \), \( a \) is one of the symbols shown in the first column of the matrix and \( b \) one of those occurring in its first line.

Given two OPMs \( M_1 \) and \( M_2 \), we define set inclusion and union:

- \( M_1 \subseteq M_2 \) if \( \forall a, b : (M_1)_{ab} \subseteq (M_2)_{ab}, \)
- \( M = M_1 \cup M_2 \) if \( \forall a, b : M_{ab} = (M_1)_{ab} \cup (M_2)_{ab} \).
Definition 1.1 (OPG and OPL). An OPG is an OPL if and only if \( M = OPM(G) \) is a conflict-free matrix, i.e., \( \forall a, b, |M_{ab}| \leq 1 \). An OPL is a language generated by an OPG.

From the above definition it is immediate to verify that both grammars of Figure 1 and of Figure 2 are OPGs (with different OPMs).

Two matrices are compatible if their union is conflict-free. A matrix is total (or complete) if it contains no empty case. The following definition of Fischer normal form (FNF) is adapted from the original one [22] to take into account that in our basic definition of CF grammar \( S \) is a set rather than a singleton.

Definition 1.2 (FNF). An OPG is in FNF if it is invertible, has no empty rule except possibly \( A \rightarrow \varepsilon \), where \( A \) is an axiom not used elsewhere, and no renaming rules.

Let \( G = (N, \Sigma, P, S) \) be an OPG; then an equivalent OPG \( \tilde{G} = (\tilde{N}, \Sigma, \tilde{P}, \tilde{S}) \) in FNF, can be built such that \( \tilde{N} = \mathcal{P}(N) \) and \( |\tilde{P}| = O(|P| \cdot 2^{|N|/2}) \), where \( k \) is the maximum length of \( P \)'s r.h.s.'s [25].

An FNF (manually) derived from the grammar of Figure 1 is given below. Notice that in this case the size of the nonterminal alphabet and of the productions is much smaller than the worst case upper bound provided by the general construction.

\[
E \rightarrow E + T \mid E + F \mid T + T \mid F + F \mid F + T \mid T + F, \\
T \rightarrow T \times F \mid F \times F, \\
F \rightarrow n \mid \langle E \rangle \mid \langle T \rangle \mid \langle F \rangle.
\]

It is well known that OPLs are a proper subfamily of deterministic CF languages: for instance, it is impossible to generate the language \( \{ a^n b a^n \} \), without producing a precedence conflict \( a < a \) and \( a > a \). Despite this theoretical limitation, OPLs have been successfully used to formalize many programming languages and to support their compilers; in this paper we will also provide several other examples of potential applications of this model in different fields.

OPMs play a fundamental role in deterministic parsing of OPLs. Thus in the view of defining automata to parse OPLs (OPAs) we pair them with the alphabet. To this goal, we use a special symbol \( \# \) not in \( \Sigma \) to mark the beginning and the end of any string. This is consistent with the typical operator parsing technique which requires the lookback and lookahead of one character to determine the precedence relation [24]. The precedence relations in the OPM are implicitly extended to include \( \# \): the initial \( \# \) can only yield precedence, and other symbols can only take precedence over the ending \( \# \).

Definition 1.3 (OP alphabet). An OP alphabet is a pair \((\Sigma, M)\), where \( \Sigma \) is an alphabet and \( M \) is a conflict-free OPM, i.e., a \(|\Sigma \cup \{\#\}|^2\) array that associates at most one of the \( \langle \rangle \) relations, =, <, or >, with each ordered pair \((a, b)\).

If \( M_{ab} = \{\circ\} \) with \( \circ \in \{=, <, >\} \), we write \( a \circ b \). For \( u, v \in \Sigma^* \) we write \( u \circ v \) if \( u = xa \) and \( v = by \) with \( a \circ b \). The relations involving the \( \# \) delimiter are constrained as stated above.

The notion of chain introduced by the following definitions provides a formal description of the intuitive concept of “invisible or hidden structure” discussed in Example 1.

Definition 1.4 (chains). Let \((\Sigma, M)\) be a precedence alphabet.

- A simple chain is a word \( a_0 a_1 a_2 \ldots a_n a_{n+1} \), written as \( a_0[a_1 a_2 \ldots a_n]a_{n+1} \), such that: \( a_0, a_{n+1} \in \Sigma \cup \{\#\}, a_i \in \Sigma \) for every \( i : 1 \leq i \leq n \), \( M_{a_i a_{i+1}} \neq \emptyset \), and \( a_0 < a_1 \leq a_2 \ldots \leq a_{n-1} \leq a_n > a_{n+1} \).
Miff the two following conditions hold:

Example 1, where a simple chain, and either \( x_i = \varepsilon \) or \( \alpha^i[x_i]^{a_{i+1}} \) is a chain (simple or composed) for every \( i : 0 \leq i \leq n \). Such a composed chain will be written as \( \alpha^0[x_0a_1x_1a_2 \ldots a_nx_n]^{a_{n+1}} \).

The body of a chain \( \alpha^i[x]^b \), simple or composed, is the word \( x \).

Example 2. The “hidden” structure induced by the OP alphabet of Example 1 for the expression \( \#n + n \times (n + n)\# \) is represented in Figure 3, where \( \#[x_0 + x_1]\# \), \( +[y_0 \times y_1]\# \), \( \times[\{w_0\}]\# \), \( \{[z_0 + z_1]\} \) are composed chains and \( \#[n]^+, +[n]^+ \) are simple chains.

**Definition 1.5 (depth of a chain).** Given a chain \( \alpha^i[x]^b \) the depth \( d(x) \) of its body \( x \) is defined recursively: \( d(x) = 1 \) if the chain is simple, whereas

\[
d(x_0a_1x_1 \ldots a_nx_n) = 1 + \max_i d(x_i).
\]

The depth of a chain is the depth of its body.

For instance, the composed chain \( \#[x_0 + x_1]\# \) in Example 2 has depth 5. Thus, if for an OPG \( G \) it is \( OPM(G) = M \), the depth of a chain body \( x \) is the height of the syntax tree, if any, whose frontier is \( x \).

**Definition 1.6 (compatible word).** A word \( w \) over \( (\Sigma, M) \) is compatible with \( M \) iff the following conditions hold:

- for each pair of letters \( c, d \), consecutive in \( w \), \( M_{cd} \neq \emptyset \);
- for each factor (substring) \( x \) of \( \#w\# \) such that \( x = a_0x_0a_1x_1a_2 \ldots a_nx_na_{n+1} \),

\[
a_0 < a_1 \leq a_2 \ldots a_{n-1} \leq a_n > a_{n+1}
\]

and, for every \( 0 \leq i \leq n \), either \( x_i = \varepsilon \) or \( \alpha^i[x_i]^{a_{i+1}} \) is a chain (simple or composed), then \( M_{a_0a_{n+1}} \neq \emptyset \).

For instance, the word \( n + n \times (n + n) \) is compatible with the OP alphabet of Example 1, whereas \( n + n \times (n + n)(n + n) \) is not.

The chains fully determine the structure of the words; in particular, given an OP alphabet, each word in \( \Sigma^* \) compatible with \( M \) is assigned a tree structure by the OPM \( M \). If \( M \) is complete, then each word is compatible with \( M \) and the OPM \( M \) assigns a structure to any word in \( \Sigma^* \). For this reason we say that OPLs somewhat generalize the notion of IDL, since their parsing is driven by the OPM which is defined on the terminal alphabet, but they also allow us to generate sentences whose structure is “invisible” before parsing.

The equal in precedence relations of an OP alphabet are connected with an important parameter of the grammar, namely, the length of the r.h.s.’s of the rules. Clearly, a rule \( A \rightarrow A_1a_1 \ldots A_ta_tA_{t+1} \), where each \( A_i \) is a possibly missing nonterminal, is
associated with relations $a_1 \doteq a_2 \doteq \cdots \doteq a_n$. If the $\doteq$ relation is cyclic, i.e., there exist $a_1, a_2, \ldots, a_n \in \Sigma \ (n \geq 1)$ such that $a_1 \doteq a_2 \doteq \cdots \doteq a_n \doteq a_1$, there is a priori no finite bound on the length of the r.h.s.’s of a production. Otherwise the length is bounded by $2 \cdot c + 1$, where $c \geq 1$ is the length of the longest $\doteq$-chain.

Previous literature \cite{18, 28} assumed that all precedence matrices of OPLs are $\doteq$-acyclic. In the case of OPGs this prevents the risk of r.h.s.’s of unbounded length \cite{19}, but could be replaced by the weaker restriction of production’s r.h.s. of bounded length, or could be removed by allowing such unbounded forms of grammars, e.g., with regular expressions as r.h.s. In our experience, such assumption helps to simplify notation and some technicalities of proofs; moreover, we found that its impact in practical examples is minimal. In this paper we accept a minimal loss of generation power and assume the simplifying assumption of $\doteq$-acyclicity. We will see, however, that this hypothesis has an impact only on constructions involving grammars but is irrelevant for the OP automata defined in this paper.

1.1. Previous results. Herein, we present some basic properties of OPLs that have already been stated in previous literature. Preliminarily, notice that, since the union of two acyclic OPMs might be cyclic, when we consider, in what follows, the union $M = M_1 \cup M_2$ of two OPMs $M_1$ and $M_2$ we always assume that $M$ too is acyclic.

**Statement 1.1** (see \cite{19}). OPLs are closed w.r.t. Boolean operations. Precisely, given two OPLs $L_1, L_2$ with compatible OPMs $M_1$ and $M_2$, $L_1 \cap L_2$ and $L_1 \cup L_2$ are OPLs whose OPM is contained in $M_1 \cup M_2$; furthermore, let $L_1^{\text{max}}$ be the OPL of all strings compatible with $M_1$, then $L_1^{\text{max}} \setminus L_1$ is an OPL whose OPM is contained in $M_1$. In particular, if $M_1$ is a complete OPM, $L_1^{\text{max}}$ is $\Sigma^*$ (where each sentence has a structure determined by $M_1$), then $\Sigma^* \setminus L_1$ is an OPL whose OPM is contained in $M_1$.

**Statement 1.2** (see \cite{18}). OPLs are closed w.r.t. concatenation and Kleene * operation. Precisely, given two OPLs $L_1, L_2$ with compatible OPMs $M_1$ and $M_2$, $L_1 \cdot L_2$ and $L_1^*$ are OPLs whose OPM is compatible with $M_1 \cup M_2$ (resp., $M_1$). Notice that in this case the construction of the new grammars may introduce new precedence relations not existing in the original matrices. Furthermore, OPLs are closed under alphabetical homomorphisms that preserve conflict freedom.

**Statement 1.3** (see \cite{18}). OPLs strictly include the family of VPLs. Precisely, VPLs are the subfamily of OPLs whose OPM is a partitioned matrix, i.e., a matrix whose structure is depicted in Figure 4.

2. OPAs. Next, we introduce a family of pushdown automata that recognize exactly OPLs. OPLs being naturally oriented towards bottom-up parsing, their accepting automata exhibit a typical shift-reduce attitude; they are considerably simpler, however, than other classical automata of this type such as LR ones.

**Definition 2.1** (OPA). A nondeterministic OPA is given by a tuple: $A = \langle \Sigma, M, Q, I, F, \delta \rangle$, where

- $(\Sigma, M)$ is an OP alphabet,
- $Q$ is a set of states (disjoint from $\Sigma$),
- $I \subseteq Q$ is a set of initial states,
- $F \subseteq Q$ is a set of final states.

---

*An example language that cannot be generated with an $\doteq$-acyclic OPM is the following: $L = \{a^n(bc)^n \mid n \geq 0\} \cup \{b^n(ca)^n \mid n \geq 0\} \cup \{a^n(ab)^n \mid n \geq 0\}$. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
\[ \delta : Q \times (\Sigma \cup Q) \to \mathcal{P}(Q) \text{ is the transition function, which is the union of three functions,} \]

\[ \delta_{\text{shift}} : Q \times \Sigma \to \mathcal{P}(Q), \quad \delta_{\text{push}} : Q \times \Sigma \to \mathcal{P}(Q), \quad \delta_{\text{pop}} : Q \times Q \to \mathcal{P}(Q). \]

We represent a nondeterministic OPA by a graph with \( Q \) as the set of vertices and \( \Sigma \cup Q \) as the set of edge labelings. The edges of the graph are denoted by different shapes of arrows to distinguish the three types of transitions: there is an edge from state \( q \) to state \( p \) labeled by \( a \in \Sigma \) denoted by a dashed (resp., normal) arrow iff \( p \in \delta_{\text{shift}}(q, a) \) (resp., \( p \in \delta_{\text{push}}(q, a) \)) and there is an edge from state \( q \) to state \( p \) labeled by \( r \in Q \) and denoted by a double arrow iff \( p \in \delta_{\text{pop}}(q, r) \).

To define the semantics of the automaton, we introduce some notation. We use letters \( p, q, x, r \) to denote states in \( Q \). Let \( \Gamma \) be \( \Sigma \times Q \) and let \( \Gamma' \) be \( \Gamma \cup \{ \perp \} \); we denote symbols in \( \Gamma' \) as \( [a, q] \) or \( \perp \). We set \( \text{symbol}([a, q]) = a \), \( \text{symbol}(\perp) = \# \), and \( \text{state}([a, q]) = q \). Given a string \( \Pi = \perp \pi_1 \pi_2 \ldots \pi_n \) with \( \pi_i \in \Gamma \), \( n \geq 0 \), we set \( \text{symbol}(\Pi) = \text{symbol}(\pi_n) \), including the particular case \( \text{symbol}(\perp) = \# \).

A configuration of an OPA is a triple \( C = \langle \Pi, q, w \rangle \), where \( \Pi \in \perp \Gamma^* \), \( q \in Q \), and \( w \in \Sigma^* \perp \). The first component represents the contents of the stack, the second component represents the current state of the automaton, while the third component is the part of input still to be read.

A computation or run of the automaton is a finite sequence of moves or transitions \( C_1 \vdash C_2 \); there are three kinds of moves, depending on the precedence relation between the symbol on top of the stack and the next symbol to read:

- **push move**: if \( \text{symbol}(\Pi) \preceq a \) then \( \langle \Pi, p, ax \rangle \vdash \langle \Pi[a, p], q, x \rangle \) with \( q \in \delta_{\text{push}}(p, a) \);
- **shift move**: if \( a \triangleleft b \) then \( \langle \Pi[a, p], q, bx \rangle \vdash \langle \Pi[b, p], r, x \rangle \) with \( r \in \delta_{\text{shift}}(q, b) \);
- **pop move**: if \( a \triangleright b \) then \( \langle \Pi[a, p], q, bx \rangle \vdash \langle \Pi, r, bx \rangle \) with \( r \in \delta_{\text{pop}}(q, p) \).

Notice that shift and pop moves are never performed when the stack contains only \( \perp \).

Push and shift moves update the current state of the automaton according to the transition functions \( \delta_{\text{push}} \) and \( \delta_{\text{shift}} \), respectively; push moves put a new element on the top of the stack consisting of the input symbol together with the current state of the automaton, whereas shift moves update the top element of the stack by changing its input symbol only. The pop move removes the symbol on the top of the stack, and the state of the automaton is updated by \( \delta_{\text{pop}} \) on the basis of the pair of states consisting of the current state of the automaton and the state of the removed stack symbol; notice that in this move the input symbol is used only to establish the \( \triangleright \) relation and it remains available for the following move.

---

**Legend**

\[\begin{array}{c|c|c|c}
\Sigma_c & \Sigma_r & \Sigma_i \\
\hline
\ll & \neq & \ll \\
\gg & \gg & \gg \\
\end{array}\]

\(\Sigma_c\) denotes “calls”

i.e., a generalized version of open parentheses;

\(\Sigma_r\) denotes “returns”

i.e., a generalized version of closed parentheses;

\(\Sigma_i\) denotes internal characters

i.e., characters that are not pushed onto the stack and are managed exclusively by finite-state control.

---

**Fig. 4** A partitioned matrix, where \( \Sigma_c, \Sigma_r, \Sigma_i \) are a set of terminal character. A precedence relation in position \( \Sigma_c, \Sigma_r, \Sigma_i \) means that relation holds between all symbols of \( \Sigma_c \) and all those of \( \Sigma_i \).
We say that a configuration $\langle \bot, q_I, x\# \rangle$ is initial if $q_I \in I$ and a configuration $\langle \bot, q_F, \# \rangle$ is accepting if $q_F \in F$. The language accepted by the automaton is defined as

$$L(A) = \left\{ x \mid \langle \bot, q_I, x\# \rangle \Rightarrow^* \langle \bot, q_F, \# \rangle, q_I \in I, q_F \in F \right\}.$$  

**Example 3.** The OPA depicted in Figure 5 accepts the language of arithmetic expressions generated by the OPG of Example 1. The same figure also shows an accepting computation on input $n + n \times (n + n)$.

Therefore, an OPA selects an appropriate subset within the “universe” of strings in $\Sigma^*$ compatible with $M$. This property somewhat resembles the fundamental Chomsky–Schützenberger theorem, in that a universe of nested structures—a Dyck language—is restricted by means of an “intersection” with a finite-state mechanism. For instance, the automaton of Figure 5 recognizes well-nested parenthesized arithmetic expressions and could be modified in such a way that parentheses are used only when needed to give the expression the desired meaning, i.e., a pair of parentheses

<table>
<thead>
<tr>
<th>stack</th>
<th>state</th>
<th>current input</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bot$</td>
<td>$q_0$</td>
<td>$n + n \times (n + n)#$</td>
</tr>
<tr>
<td>$\bot, n, q_0$</td>
<td>$q_1$</td>
<td>$n + n \times (n + n)#$</td>
</tr>
<tr>
<td>$\bot$</td>
<td>$q_1$</td>
<td>$n \times (n + n)#$</td>
</tr>
<tr>
<td>$\bot, +, q_1$</td>
<td>$q_0$</td>
<td>$n \times (n + n)#$</td>
</tr>
<tr>
<td>$\bot, +, q_1, n, q_0$</td>
<td>$q_1$</td>
<td>$(n + n)#$</td>
</tr>
<tr>
<td>$\bot, +, q_1 \times, q_1 #$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bot, +, q_1, \times, q_1 (+, q_0$</td>
<td>$q_2$</td>
<td>$n \times (n + n)#$</td>
</tr>
<tr>
<td>$\bot, +, q_1, \times, q_1 [n, q_2$</td>
<td>$q_3$</td>
<td>$+n#$</td>
</tr>
<tr>
<td>$\bot, +, q_1, \times, q_1 [q_0 + q_3$</td>
<td>$q_3$</td>
<td>$+n#$</td>
</tr>
<tr>
<td>$\bot, +, q_1, \times, q_1 [q_0 +, q_3$</td>
<td>$q_3$</td>
<td>$+n#$</td>
</tr>
<tr>
<td>$\bot, +, q_1, \times, q_1 [q_0$</td>
<td>$q_3$</td>
<td>$#$</td>
</tr>
<tr>
<td>$\bot, +, q_1$</td>
<td>$q_3$</td>
<td>$#$</td>
</tr>
<tr>
<td>$\bot$</td>
<td>$q_3$</td>
<td>$#$</td>
</tr>
</tbody>
</table>

**Fig. 5.** Automaton and example of computation for the language of Example 3. Recall that shift, push, and pop transitions are denoted by dashed, normal, and double arrows, respectively.
containing a + is necessary only if it is adjacent to a ×; parentheses enclosing only × should be avoided.

The following definitions will be used throughout the paper to characterize OPA behavior: we use arrows $\rightarrow$, $\Rightarrow$, and $\impliedby$ to denote push, shift, and pop transitions, respectively.

**Definition 2.2.** Let $A$ be an OPA. A support for a simple chain $a_0[a_1a_2\ldots a_n]^{a_{n+1}}$ is any path in $A$ of the form

$$q_0 \xrightarrow{a_1} q_1 \rightarrow \cdots \rightarrow q_{n-1} \xrightarrow{a_n} q_n \xrightarrow{q_0} q_{n+1}. \tag{2.1}$$

Notice that the label of the last (and only) pop is exactly $q_0$, i.e., the first state of the path; this pop is executed because of relations $a_0 < a_1$ and $a_n > a_{n+1}$.

A support for the composed chain $a_0[x_0a_1x_1a_2\ldots a_nx_n]^{a_{n+1}}$ is any path in $A$ of the form

$$q_0 \xrightarrow{x_0} q_0 \xrightarrow{a_1} q_1 \xrightarrow{x_1} q'_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} q_n \xrightarrow{x_n} q'_n \xrightarrow{q_0} q_{n+1}, \tag{2.2}$$

where for every $i: 0 \leq i \leq n$,

- if $x_i \neq \varepsilon$, then $q_i \xrightarrow{x_i} q'_i$ is a support for the (simple or composed) chain $a_i[x_i]^{a_{i+1}}$;
- if $x_i = \varepsilon$, then $q'_i = q_i$.

Notice that the label of the last pop is exactly $q_0$.

The support of a chain with body $x$ will be denoted by $q_0 \xrightarrow{x} q_{n+1}$.

**Example 4.** Figure 6 illustrates the supports of the chains that, for the OPA described in Example 3, compose the structure of the expression $n + n \times (n + n)$.

The chains fully determine the structure of the computation of any automaton on a word compatible with $M$. Indeed, let $\Pi \in \varDelta \Gamma^*$ with symbol($\Pi$) = $a < x > b$: an OPA $A$ performs the computation $\langle \Pi, q, xb \rangle \Rightarrow \langle \Pi, p, b \rangle$ without changing the portion $\Pi$ of the stack, iff $a[x]^b$ is a chain over $(\Sigma, M)$ with a support $q \xrightarrow{x} p$ in $A$.

The depth of $x$ corresponds to the maximum number of push/pop pairs nested in the computation, i.e., the maximum height reached by the stack in one of the traversed configurations, minus the height of the stack in the starting configuration.

Notice that the context $a, b$ of a chain is used by the automaton to build its support only because $a < x$ and $x > b$; thus, the chain's body contains all information needed by the automaton to build the subtree whose frontier is that string, once it is understood that its first move is a push and its last one is a pop. This is a distinguishing feature of OPLs, not shared by other deterministic languages: we call

Fig. 6. Structure of chains and supports for the expression of Example 4.
it the locality principle of OPLs, which is exploited elsewhere, e.g., to build parallel and/or incremental OP parsers [7].

With reference to Example 3 and Figure 6, the parsing of substring \( n + n \) within the context \( \langle \rangle \) is given by the computation

\[
\langle \Pi, q_2, n + n \# \rangle \xrightarrow{*} \langle \Pi, q_3, \# \rangle \quad \text{with} \quad \Pi = \perp + [\times, q_1] (\ell, q_0)
\]

which corresponds to support \( q_2 \xrightarrow{n} q_3 \xrightarrow{+} q_2 \xrightarrow{n} q_3 \xrightarrow{\sigma} q_3 \) of the composed chain \( \ell [n + n] \), where \( q_2 \xrightarrow{n} q_3 \) is the support \( q_2 \xrightarrow{n} q_3 \xrightarrow{\sigma} q_3 \) of the simple chains \( \ell [n] + \) and \( + [n] \).

2.1. Examples. In this section we illustrate an example of application of OPLs, which cannot be modeled by traditional classes of languages with an “explicit” structure such as parenthesis languages and VPLs. We shall present in section 4.3 examples in other interesting contexts (such as operating systems) which can be naturally modeled by OPAs recognizing \( \omega \)-languages, and are not recognizable by VPAs as well. Other examples of application of OPLs to model systems in various application fields outside the traditional one of programming languages are given in [35].

Indeed, the most distinguishable feature of the structure of VPLs is that in their OPMs the \( \div \) relation occurs always and only between open and closed parentheses (\( \Sigma_c \) and \( \Sigma_r \) elements in the notation of [3], resp.). Unlike traditional parenthesis languages, however, in VPLs parentheses can remain unmatched, but only at the beginning (\( \Sigma_c \) elements) and end (\( \Sigma_r \) elements) of the input string, respectively. This initial extension, however, is not sufficiently general to cover several interesting cases where an “event” of special type, e.g., a rollback or an exception, should force flushing the stack of many pending elements, say write operations or procedure calls.

Example 5. OPAs can be used to model the run-time behavior of database systems, e.g., for modeling sequences of users’ transactions with possible rollbacks. Other systems that exhibit an analogous behavior are revision control (or versioning) systems (such as subversion or git). As an example, consider a system for version management of files where a user can perform the following operations on documents: save them, access and modify them, undo one (or more) previous changes, restoring the previously saved version.

The following alphabet represents the user’s actions: \( sv \) (for save), \( wr \) (for write, i.e., the document is opened and modified), \( ud \) (for a single undo operation), \( rb \) (for a rollback operation, where all the changes occurred since the previously saved version are discarded).

An OPA that models the traces of possible actions of the user on a given document is a single-state automaton \( \langle \Sigma, M, \{ q \}, \{ q \}, \delta \rangle \), where \( \Sigma = \{ sv, rb, wr, ud \} \), \( M \) is

\[
\begin{array}{c|cccc}
sv & rb & wr & ud & \#
\hline
sv & < & \neq & < & >
rb & > & > & > & >
wr & < & > & < & \neq & >
ud & > & > & > & >
\# & < & < & < & \neq
\end{array}
\]

and \( \delta_{\text{push}}(q, a) = q \quad \forall a \in \{ sv, wr \} \), \( \delta_{\text{shift}}(q, a) = q \quad \forall a \in \{ rb, ud \} \) and \( \delta_{\text{pop}}(q, q) = q \).

A more specialized model of this system might impose that the user regularly backs her work up, so that no more than \( N \) changes that are not undone (denoted \( wr \) as before) can occur between any two consecutive checkpoints \( sv \) (without any
rollback \( rb \) between them). Figure 7 shows the corresponding OPA with \( N = 2 \), with the same OPM \( M \). States 0, 1 and 2 denote, respectively, the presence of zero, one, and two unmatched changes between two symbols \( sv \).

An example of computation on the string \( sv \ wr \ ud \ rb \ sv \ wr \ ud \ sv \ wr \ rb \ wr \ sv \) is shown in Table 2.

### 2.2. Determinism vs. nondeterminism

An important property of OPAs is the equivalence between the deterministic and the nondeterministic version thereof. This result also implies the closure of OPLs under complementation, yielding an alternative proof to the traditional one presented in [19].

The deterministic version of OPAs is defined along the usual lines.

**Definition 2.3 (deterministic OPA).** An OPA is deterministic if \( I \) is a singleton, and the ranges of \( \delta_{\text{push}} \), \( \delta_{\text{shift}} \), and \( \delta_{\text{pop}} \) are \( Q \) rather than \( \wp(Q) \).

It is well known that the equivalence between nondeterministic and deterministic machines usually does not extend from finite-state to pushdown ones. VPAs are however a noticeable exception. The construction described in [4] is extended here to cover OPAs too. Our construction ensures that two different pop moves of two different runs of the nondeterministic automaton never “mix up” their initial and final states in the deterministic one by keeping track of the path of the automaton since the push move that marks the origin of the chain to be reduced by the next pop move. Precisely, the states of the deterministic automaton \( \hat{A} \) are sets of pairs of states, instead of sets of single states, of the nondeterministic automaton \( A \): \( \hat{A} \) simulates \( A \) along the first component of the pair, whereas the second component stores the state that gave origin to a push transition and it is propagated through shift moves. The deterministic pop operations will simulate only the nondeterministic ones defined on the states corresponding to the first component of the current state and the state reached before the last push move, which corresponds to the state on the top of the stack in an actual run of the nondeterministic automaton.

The following theorem formalizes the above informal reasoning.

**Theorem 2.4.** Given a nondeterministic OPA \( A \) with \( s \) states, an equivalent deterministic OPA \( \hat{A} \) can effectively be built with \( 2^{O(s^2)} \) states.
Table 2
Example of computation for the specialized system of Example 5.

<table>
<thead>
<tr>
<th>stack</th>
<th>state</th>
<th>current input</th>
</tr>
</thead>
<tbody>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud rb sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud rb sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₄</td>
<td>℘ud rb sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₄</td>
<td>℘vb sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₁</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₁</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₄</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₄</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₁</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₄</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₄</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₁</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
<tr>
<td>⊥</td>
<td>q₀</td>
<td>℘sv wr ud sv wr rb wr sv #</td>
</tr>
</tbody>
</table>

Proof. Let $A$ be $(\Sigma, M, Q, I, F, \delta)$; $\hat{A} = (\Sigma, M, \hat{Q}, I, F, \hat{\delta})$ is defined as follows:

- $\hat{Q} = P(Q \times (Q \cup \{\top\}),$ where $Q \cap \{\top\} = \emptyset$ and $\top$ is a symbol that stands for the baseline of the computations; we will use $K, K_1, K', \ldots$ to denote states in $\hat{Q}$,
- $I = I \times \{\top\}$ is the initial state of $\hat{A}$,
- $\hat{F} = \{K \mid K \cap (F \times \{\top\}) \neq \emptyset\}$,
- $\hat{\delta} : Q \times (\Sigma \cup \hat{Q}) \to \hat{Q}$ is the transition function defined as follows.

The push transition $\delta_{\text{push}} : Q \times \Sigma \to \hat{Q}$ is defined by

$$\delta_{\text{push}}(K, a) = \bigcup_{(q,p) \in K} \{(h,q) \mid h \in \delta_{\text{push}}(q,a)\}.$$ 

The shift transition $\delta_{\text{shift}} : Q \times \Sigma \to \hat{Q}$ is defined by

$$\delta_{\text{shift}}(K, a) = \bigcup_{(q,p) \in K} \{(h,p) \mid h \in \delta_{\text{shift}}(q,a)\}.$$ 

The pop transition $\delta_{\text{pop}} : \hat{Q} \times Q \to \hat{Q}$ is defined as follows:

$$\delta_{\text{pop}}(K_1, K_2) = \bigcup_{(r,q) \in K_1, (q,p) \in K_2} \{(h,p) \mid h \in \delta_{\text{pop}}(r,q)\}.$$ 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Notice that, if $|Q| = s$ is the number of states of the nondeterministic OPA $A$, the deterministic OPA $\tilde{A}$ that is obtained in this way has a set of states whose size is exponential in $s^2$, i.e., $|\tilde{Q}| = 2^{Q:|Q酌|}$ which is $2^O(s^2)$.

The proof of equivalence between the two automata is by induction and is based on Lemmas 2.5 and 2.6.

**Lemma 2.5.** Let $y$ be the body of a chain with support $q \not\Rightarrow y$ in $\tilde{A}$. Then, for every $p \in Q$ and $K \in \tilde{Q}$, if $K \ni (q, p)$, there exists a support $K' \ni y$ in $\tilde{A}$ with $K' \ni (q', p)$.

Proof. We argue by induction on the depth $h$ of $y$. If $h = 1$ then $y = a_1a_2 \ldots a_n$ and the support can be rewritten as in (2.1) with $q_0 = q$ and $q_{n+1} = q'$. Set $K_0 = K'$ and

$$
K_1 = \delta_{\text{push}}(K_0, a_1),
K_i = \delta_{\text{shift}}(K_{i-1}, a_i), \text{ for every } i = 2, \ldots, n,
K' = \delta_{\text{pop}}(K_n, K).
$$

Then

$$
(2.3)
$$

is a support for $C$ in $\tilde{A}$. Moreover, since $K \ni (q, p)$, by the definition of $\delta$ we have

$$
K_1 \ni (q_1, q) \quad \text{since } \delta_{\text{push}}(q, a_1) \ni q_1,
K_i \ni (q_i, q) \quad \text{since } \delta_{\text{shift}}(q_{i-1}, a_i) \ni q_i,
K' \ni (q', p) \quad \text{since } \delta_{\text{pop}}(q_n, q) \ni q'.
$$

Now assume that the statement holds for supports with depth lower than $h$ and let $y = x_0a_1x_1a_2 \ldots a_n x_n$ have depth $h$. The support can be rewritten as in (2.2) with $q_0 = q$ and $q_{n+1} = q'$, where $q_i = q_i$ whenever $x_i$ is the empty word, and every nonempty $x_i$ has depth lower than $h$.

Then, by the inductive hypothesis and the definition of $\delta$, we can build a support

$$
(2.4)
$$

where, $(q, p)$ being in $K$, we have

$$
K_0' \ni (q_0', p) \quad \text{by inductive hypothesis on the support } q = q_0 \not\Rightarrow q_0',
K_1 \ni (q_1, q_0') \quad \text{since } \delta_{\text{push}}(q_0', a_1) \ni q_1,
K_i' \ni (q_i', q_0') \quad \text{by inductive hypothesis on the support } q_i \not\Rightarrow q_i',
K_n' \ni (q_n, q_0') \quad \text{since } \delta_{\text{shift}}(q_{i-1}, a_i) \ni q_i \text{ for every } i = 2, \ldots, n,
K' \ni (q', p) \quad \text{by inductive hypothesis on the support } q \not\Rightarrow q',
$$

and this concludes the proof.

**Lemma 2.6.** Let $y$ be the body of a chain with support $K \not\Rightarrow K'$ in $\tilde{A}$. Then, for every $p, q' \in Q$, if $K' \ni (q', p)$ there exists a support $q \not\Rightarrow q'$ in $\tilde{A}$ with $(q, p) \in K$.

Proof. First we present some remarks we will use in the proof.

(i) By the definition of $\delta_{\text{push}}$, if $K \not\Rightarrow K'$ in $\tilde{A}$, $(\tilde{q}, q) \in K$, $(q, p) \in K$, then $q \not\Rightarrow \tilde{q}$ in $\tilde{A}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
(ii) By the definition of \( \delta_{\text{shift}} \), if \( K \xrightarrow{a} K \) in \( \tilde{A} \), \((r,q) \in K\), then there exists a state \( \tilde{q} \in Q \) such that \( \tilde{q} \xrightarrow{a} r \) in \( A \) and \((\tilde{q},q) \in \tilde{K}\).

(iii) By the definition of \( \delta_{\text{pop}} \), if \( \tilde{K} \xrightarrow{q} K' \) in \( \tilde{A} \) and \((q',p) \in K'\), then there exists a pair \((r,q) \in \tilde{K}\) such that \((q,p) \in K \) and \( r \xrightarrow{q} q' \) in \( A \).

We argue by induction on the depth \( h \) of \( y \). If \( h = 1 \), then \( y = a_1a_2\ldots a_n \) and the support can be rewritten as in (2.3). Let \( K' \ni (q',p); \) then, by remark (3) there exists a pair \((q_n,q) \in K_n \) such that \((q,p) \in K \) and \( q_n \xrightarrow{q} q' \) in \( \tilde{A} \). Moreover, \((q_n,q) \in K_n \) and \( K_n-1 \xrightarrow{a_{n-1}} K_n \) imply by remark (2) the existence of a state \( q_{n-1} \in Q \) such that \((q_{n-1},q) \in K_{n-1} \) and \( q_{n-1} \xrightarrow{a_n} q_n \). Similarly one can verify that for every \( i = n-2,\ldots,1 \) there exists \( q_i \in Q \) such that \((q_i,q) \in K_i \) and \( q_i \xrightarrow{a_{i+1}} q_{i+1} \). Finally, \( K \xrightarrow{a_i} K_1 \), \((q_1,q) \in K_1 \), and \((q,p) \in K \) imply by remark (1) that \( q \xrightarrow{a} q_1 \) in \( A \). Thus, we built backward a path as in (2.1) with \( q_0 = q \), \( q_{n+1} = q' \), \((q,p) \in K \), and this concludes the proof of the induction basis.

Now assume that the statement holds for chains with depth lower than \( h \). Let \( y = x_0a_1x_1a_2\ldots a_nx_n \) have depth \( h \) and consider a support as in (2.4) where \( K' = K_i \) whenever \( x_i \) is the empty word, and every nonempty \( x_i \) has depth lower than \( h \). Let \((q',p) \in K' \). Since \( K_n \xrightarrow{q_i} K_1 \), by remark (3) there exists a pair \((q_n',q_0') \in K_n' \) with \((q_0',p) \in K_0' \) and \( q_n' \xrightarrow{q} q' \) in \( \tilde{A} \). If \( x_n \neq \varepsilon \), by the inductive hypothesis, since \((q_n',q_0') \in K_n' \) there exists a support \( q_n \xrightarrow{x_n} q_n' \) with \((q_n,q_0') \in K_n \).

Similarly one can see that, for all \( i = n-1,\ldots,2,1 \), there exist \( q_i' \) and \( q_i \) \((q_i' = q_i \) whenever \( x_i \) is empty) such that

\[ q_i \xrightarrow{x_i} q_i' \xrightarrow{a_{i+1}} q_{i+1} \]

with \((q_i,q_0') \in K_i' \) by remark (2) (since \( K_i' \xrightarrow{a_{i+1}} K_{i+1} \) in \( \tilde{A} \) and \((q_{i+1},q_0') \in K_{i+1} \)), and \((q_i',q_0') \in K_i \) by the inductive hypothesis (since \( K_i \xrightarrow{x_i} K_i' \) in \( \tilde{A} \) and \((q_i',q_0') \in K_i' \)).

In particular \( q_1 \xrightarrow{x_1} q_1' \) with \((q_1,q_0') \in K_1 \). Then, since also \( K_0' \xrightarrow{a} K_1 \) and \((q_0',p) \in K_0', \) by remark (1) we get \( q_0' \xrightarrow{a} q_1 \). Finally, since \((q_0',p) \in K_0' \) and \( K \xrightarrow{q_0} K_0' \), if \( x_0 \neq \varepsilon \) the inductive hypothesis implies the existence of a state \( q \in Q \) such that \( q \xrightarrow{x_0} q_0' \) in \( A \) with \((q,p) \in K \). Hence we built a support as in (2.2) with \( q_0 = q \), \( q_{n+1} = q' \), \((q,p) \in K \), and this concludes the proof.

To complete the proof of Theorem 2.4, we prove that there exists an accepting computation for \( y \) in \( A \) iff there exists an accepting computation for \( y \) in \( \tilde{A} \).

Let \( y \) be in \( L(A) \). Then it admits a support \( q \xrightarrow{y} q' \) with \( q \in I \) and \( q' \in F \). Then for \( K = I \times \{\top\} \ni (q_0,\top) \), Lemma 2.5 implies the existence of a support \( K \xrightarrow{y} K' \) in \( \tilde{A} \) with \( K' \ni (q',\top) \). \( q' \in F \) implies \( K' \in \tilde{F} \), hence \( y \) is accepted by \( A \).

Conversely, let \( y \) be in \( L(A) \). Then \( y \) admits a support \( \tilde{K} \xrightarrow{y} K' \) in \( \tilde{A} \), with \( K' \in \tilde{F} \). This means that there exists \( q' \in F \) such that \((q',\top) \in K' \). Hence, by Lemma 2.6, there exists a support \( q \xrightarrow{y} q' \) in \( A \) with \((q',\top) \in \tilde{K} \), and this implies \( q \in I \). Thus the support \( q \xrightarrow{y} q' \) defines an accepting computation for \( y \) in \( A \).

### 2.3. Complexity of OPL decision problems.

To conclude this section we point out that the basic decision problems for OPLs have the same order of complexity as those for VPLs; precisely,

- the emptiness problem is in PTIME, OPLs and VPLs being a subclass of CF languages;

---

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
• the containment problem for deterministic OPAs is in PTIME too since it is reduced to the intersection, complement, and emptiness problems which are all in PTIME in the deterministic case;
• the containment problem in the nondeterministic case is instead EXPTIME complete: the same arguments used in [4] for VPLs apply identically for OPLs.

3. OPAs and OPGs.
Our next result is the equivalence between OPGs and OPAs.

3.1. From OPGs to OPAs.
Theorem 3.1. Let \( G = \langle N, \Sigma, P, S \rangle \) be an OPG; then an OPA \( A \) such that \( L(A) = L(G) \) can effectively be built. Furthermore, let \( m \) be the sum of the lengths of the r.h.s.’s of \( G \); then \( A \) has \( O(m^2) \) states.

Proof. First, we describe a procedure to build a nondeterministic OPA \( A = \langle \Sigma, M, Q, I, F, \delta \rangle \) from a given OPG \( G \) with the same precedence matrix \( M \) as \( G \). Then we prove the equivalence between \( A \) and \( G \).

The construction sharply differs from the traditional one involving CF grammars and general pushdown automata, which is instead quite straightforward. This is due to the remarkable peculiarities of OPAs—among them the locality principle—which make them, in turn, significantly different from the more powerful general pushdown automata and from the less powerful VPAs. To keep the construction as simple as possible, we avoid introducing any optimization. Also, without loss of generality, we assume that the grammar \( G \) has no empty or renaming rules.

\( A \) is built in such a way that a successful computation thereof corresponds to building bottom-up a derivation tree in \( G \): the automaton performs a push transition when it reads the first terminal of a new r.h.s. It performs a shift transition when it reads a terminal symbol inside a r.h.s., i.e., a leaf with some left sibling leaf. It performs a pop transition when it completes the recognition of a r.h.s., then guesses (nondeterministically) the nonterminal at the l.h.s. Each state contains two pieces of information: the first component represents the prefix of the r.h.s. under construction, whereas the second component is used to recover the r.h.s. previously under construction (see Figure 8) whenever all r.h.s.’s nested below have been completed.

Precisely, the construction of \( A \) is defined as follows. Let

\[
\mathcal{P} = \{ \alpha \in (N \cup \Sigma)^* \Sigma \mid \exists A \rightarrow \alpha \beta \in P \}
\]

be the set of prefixes, ending with a terminal symbol, of r.h.s.’s of \( G \); define \( Q = \{ \varepsilon \} \cup \mathcal{P} \cup N, Q = Q \times \{ \varepsilon \} \cup \mathcal{P}, I = \{ (\varepsilon, \varepsilon) \}, \) and \( F = S \times \{ \varepsilon \} \cup \{ (\varepsilon, \varepsilon) \mid \varepsilon \in L(G) \} \).

Note that \( |Q| = 1 + |\mathcal{P}| + |N| \) is \( O(m) \); therefore \( |Q| \) is \( O(m^2) \).
The transition functions are defined as follows for \(a \in \Sigma\) and \(\alpha, \alpha_1, \alpha_2 \in \mathbb{Q}, \beta, \beta_1, \beta_2 \in \{\varepsilon\} \cup \mathbb{P}:
\[
\begin{align*}
\delta_{\text{shift}}(\langle \alpha, \beta \rangle, a) & \ni \begin{cases} 
\langle \alpha a, \beta \rangle & \text{if } \alpha \notin N, \\
\langle \beta a, \beta \rangle & \text{if } \alpha \in N,
\end{cases} \\
\delta_{\text{push}}(\langle \alpha, \beta \rangle, a) & \ni \begin{cases} 
\langle a, \alpha \rangle & \text{if } \alpha \notin N, \\
\langle \alpha a, \beta \rangle & \text{if } \alpha \in N,
\end{cases} \\
\delta_{\text{pop}}(\langle \alpha_1, \beta \rangle, \langle \alpha_2, \beta_2 \rangle) & \ni \langle A, \gamma \rangle \text{ for every } A \text{ such that} \\
A \rightarrow \alpha_1 & \in P \text{ if } \alpha_1 \notin N, \\
A \rightarrow \beta_1 \alpha_1 & \in P \text{ if } \alpha_1 \in N, \text{ and } \gamma = \begin{cases} 
\alpha_2 & \text{if } \alpha_2 \notin N, \\
\beta_2 & \text{if } \alpha_2 \in N.
\end{cases}
\end{align*}
\]
\]

Notice that the result of \(\delta_{\text{shift}}\) and \(\delta_{\text{push}}\) is a singleton, whereas \(\delta_{\text{pop}}\) may produce several states, in case of repeated r.h.s.'s.

The states reached by push and shift transitions have the first component in \(\mathbb{P}\). If state \(\langle \alpha, \beta \rangle\) is reached after a push transition, then \(\alpha\) is the prefix of the r.h.s. that is currently under construction and \(\beta\) is the prefix previously under construction; in this case \(\alpha\) is either a terminal symbol or a nonterminal followed by a terminal one. If the state is reached after a shift transition, then \(\alpha\) is the concatenation of the first component of the previous state with the read character, and \(\beta\) is not changed from the previous state. The states reached by a pop transition have the first component in \(\mathbb{N}\): if \(\langle A, \gamma \rangle\) is such a state, then \(A\) is the corresponding l.h.s., and \(\gamma\) is the prefix previously under construction.

The equivalence between \(G\) and \(A\) derives from the following Lemmas 3.2 and 3.3, when \(\beta = \gamma = \varepsilon\), \(\Pi = \bot\), and \(A\) is an axiom.

Example 6. Let \(G\) be the grammar introduced in Example 1. To apply the construction of Theorem 3.1 first we need to transform \(G\) in such a way that there are no renaming rules. The new grammar has the following productions,
\[
\begin{align*}
E & \rightarrow E + T \mid T \times F \mid n \mid \langle E \rangle, \\
T & \rightarrow T \times F \mid n \mid \langle E \rangle, \\
F & \rightarrow n \mid \langle E \rangle,
\end{align*}
\]
where \(E, T,\) and \(F\) are axioms.

Figure 9 shows an accepting computation of the equivalent automaton, together with the corresponding derivation tree. Notice that the computation shown in Figure 9 is equal to that of Figure 5 up to a renaming of the states; in fact, the shape of syntax trees and consequently the sequence of push, shift, and pop moves in OPLs depends only on the OPM, not on the visited states.

Lemma 3.2. Let \(x\) be the body of a chain and \(\beta, \gamma \in \mathbb{P} \cup \{\varepsilon\}\). Then, for all \(h \geq 1\),
\[
\langle \beta, \gamma \rangle \xrightarrow{\leq} q \text{ implies the existence of } A \in \mathbb{N} \text{ such that } A \Rightarrow x \text{ in } G \text{ and } q = \langle A, \beta \rangle.
\]

Proof. We reason by induction on the depth \(h\) of \(x\).

If \(h = 1\), then \(x = a_1 a_2 \ldots a_n\) is the body of a simple chain, and the support is as in (2.1) with \(q_0 = \langle \beta, \gamma \rangle\) and \(q_{n+1} = q\). Then by the definition of push and shift transition functions we have \(q_i = \langle a_1 \ldots a_i, \beta \rangle\) for every \(i = 1, 2, \ldots, n\), and by the definition of pop transition function (recall that \(\beta \notin N\) by hypothesis) it is \(q = \langle A, \beta \rangle\) for some \(A\) such that \(A \rightarrow a_1 \ldots a_n = x\) in \(G\). Hence \(A \Rightarrow x\) and the statement is proved.

If \(h > 1\), then as usual let \(x = x_0 a_1 x_1 \ldots a_n x_n\) and let its support be decomposed as in (2.2) with \(q_0 = \langle \beta, \gamma \rangle\) and \(q_{n+1} = q\). Also set \(q_i = \langle \beta_i, \gamma_i \rangle\) for \(i = 0, 1, \ldots, n\) (in particular \(\beta_0 = \beta\) and \(\gamma_0 = \gamma\)). Each nonempty \(x_i\) being the body of a chain with depth lower than \(h\), the inductive hypothesis implies that there exists \(X_i \in N\) such
that \( X_i \Rightarrow x_i \) in \( G \), and \( q_i = \langle X_i, \beta_i \rangle \). Thus, the support can be rewritten as

\[
\langle \beta, \gamma \rangle \xrightarrow{\alpha_0} q_0 \xrightarrow{a_1} \langle \beta_1, \gamma_1 \rangle \xrightarrow{\alpha_1} q_1 \xrightarrow{a_2} \cdots \xrightarrow{a_n} \langle \beta_n, \gamma_n \rangle \xrightarrow{\alpha_n} q_n \xrightarrow{\beta_n} q,
\]

where

\[
q_i = \begin{cases} 
\langle \beta_i, \gamma_i \rangle & \text{if } x_i = \varepsilon, \\
\langle X_i, \beta_i \rangle & \text{otherwise}
\end{cases}
\]

for every \( i \). Now, by the definition of push and shift transition functions, one can see that, for \( i \neq 0 \), \( \beta_i = X_0a_1 \cdots X_{i-1}a_i \) holds regardless of whether \( x_i \) is empty or not (setting \( X_i = \varepsilon \) if \( x_i = \varepsilon \)). Thus, to compute the state \( q \) reached with the final pop transition \( \delta_{\text{pop}}(q'_n, q'_0) \), we have to consider four cases depending on whether \( x_0 \) and \( x_n \) are empty or not, which are exactly the four combinations considered in the definition of \( \delta_{\text{pop}} \). In any case, \( q \) has the form \( \langle A, \beta \rangle \), where \( A \) is a nonterminal of \( G \) such that \( A \Rightarrow X_0a_1 X_1 \cdots X_{n-1}a_n X_n \). \( \square \)

**Lemma 3.3.** Let \( x \) be the body of a chain and \( A \in N \). Then, \( A \Rightarrow x \) in \( G \) implies \( \langle \beta, \gamma \rangle \xrightarrow{\epsilon} \langle A, \beta \rangle \) for every \( \beta, \gamma \in P \cup \{ \varepsilon \} \).

**Proof.** We reason by induction on the depth \( h \) of the chain. If \( h = 1 \), then \( x \) is the body of a simple chain, hence \( A \Rightarrow x \) means that \( A \to x \) is a production. Thus, by

<table>
<thead>
<tr>
<th>stack</th>
<th>state</th>
<th>current input</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \perp )</td>
<td>( (\varepsilon, \varepsilon) )</td>
<td>( n + n \times [n + n] # )</td>
</tr>
<tr>
<td>( \perp[n, \langle \varepsilon, \varepsilon \rangle] )</td>
<td>( (n, \varepsilon) )</td>
<td>( +n \times [n + n] # )</td>
</tr>
<tr>
<td>( \perp )</td>
<td>( (E, \varepsilon) )</td>
<td>( +n \times [n + n] # )</td>
</tr>
<tr>
<td>( \perp \times, (E, \varepsilon) )</td>
<td>( (E +, \varepsilon) )</td>
<td>( n \times [n + n] # )</td>
</tr>
<tr>
<td>( \perp \times, (E, \varepsilon) )</td>
<td>( n \times [n + n] # )</td>
<td></td>
</tr>
<tr>
<td>( \perp[+, (E, \varepsilon)] \times, (T, E+) )</td>
<td>( (T \times, E+) )</td>
<td>( (n + n) # )</td>
</tr>
<tr>
<td>( \perp[+, (E, \varepsilon)] \times, (T, E+) )</td>
<td>( (T \times, E+) )</td>
<td>( (n + n) # )</td>
</tr>
</tbody>
</table>

**FIG. 9. Accepting computation of the automaton built in Theorem 3.1.**
the definition of $\delta$ (recall that $\beta \not\in N$ by hypothesis), we obtain a support as in (2.1) with $q_0 = \langle \beta, \gamma \rangle$, $q_{n+1} = q$, and $q_i = \langle a_1 \ldots a_i, \beta \rangle$ for every $i = 1, 2, \ldots n$.

If $h > 1$, then $x$ is the body of a composed chain with $x = x_0 a_1 x_1 \ldots a_n x_n$. Hence $A \Rightarrow x$ in $G$ implies that there exist $X_0, X_1, \ldots, X_n \in \{ \varepsilon \} \cup N$ (more precisely, $X_i = \varepsilon$ if $x_i = \varepsilon$) such that $A \Rightarrow X_0 a_1 X_1 \ldots a_n X_n$ and $X_i \Rightarrow x_i$. The first step of the computation is different depending on whether $x_0$ is empty or not. In any case, we have

$$\langle \beta, \gamma \rangle \xrightarrow{a_0} q_0 \xrightarrow{a_1} \langle X_0 a_1, \beta \rangle,$$

where

$$q_0 = \begin{cases} \langle \beta, \gamma \rangle & \text{if } x_0 = \varepsilon, \\ \langle X_0, \beta \rangle & \text{otherwise.} \end{cases}$$

The computation goes on differently depending on whether $x_1, x_2, \ldots, x_{n-1}$ are empty or not. However, by the inductive hypothesis and the definition of $\delta_{\text{shift}}$, after reading $a_i$ the automaton reaches state $\langle X_0 a_1 \ldots X_{i-1} a_i, \beta \rangle$ for every $i = 1, \ldots, n$, i.e., we have the path

$$\langle \beta, \gamma \rangle \xrightarrow{a_0} q_0 \xrightarrow{a_1} \langle X_0 a_1, \beta \rangle \xrightarrow{a_2} \langle X_0 a_1 X_1 a_2, \beta \rangle \xrightarrow{a_3} \langle X_0 a_1 \ldots a_3, X_{n-1} a_n, \beta \rangle.$$ 

If $x_n \neq \varepsilon$, the computation proceeds with the last inductive step

$$\langle X_0 a_1 \ldots X_{n-1} a_n, \beta_n \rangle \xrightarrow{a_n} \langle X_n, X_0 a_1 \ldots X_{n-1} a_n \rangle.$$ 

Finally, the computation ends with a pop transition. There are four cases depending on whether $x_0$ and $x_n$ are empty or not, which are exactly the four combinations considered in the definition of $\delta_{\text{pop}}$. In any case, we build a support ending with state $\langle A, \beta \rangle$, and this concludes the proof.

**Corollary 3.4.** If the source grammar is in FNF, then the corresponding automaton is deterministic.

The thesis follows immediately by observing that the construction defined in Theorem 3.1 is such that the values defined by $\delta_{\text{push}}$ and $\delta_{\text{shift}}$ are always singleton, whereas $\delta_{\text{pop}}$ produces as many states as many l.h.s.'s have the same r.h.s.'s. Thus, since the initial state is a singleton and grammars in FNF have no repeated r.h.s.'s, the automaton resulting from the construction is already deterministic. This corollary has an interesting effect in terms of size of the produced automata as pointed out below.

**Remark 1.** Given a grammar $G$ with $|N|$ nonterminals the construction of Theorem 3.1 produces an automaton with $O(m^2)$ states, where $m$ is defined as in Theorem 3.1; thus, if we build a deterministic OPA from a generic OPG by first building a nondeterministic automaton and then transforming it into the deterministic version, we obtain an automaton with $2O(m^2)$ states; instead, if we first transform the original $G$ in FNF we obtain an equivalent grammar $\tilde{G}$ with $O(2^{|N|})$ nonterminals and $\tilde{m} = O(2^{|N|^2})$; then, by applying the construction of Theorem 3.1 we directly obtain a deterministic automaton with $O(\tilde{m}^2) = O(2^{2|N|^2})$ states.

Nevertheless, the size of the complete automaton is clearly hardly manageable by human execution; thus we implemented a prototype (nonoptimized) tool to perform the construction.\(^5\)

\(^5\)The tool is called **Flup**, available at https://github.com/bzoto/flup.
3.2. From OPAs to OPGs.

The construction of an OPG equivalent to a given OPA is far simpler than the converse one, thanks to the explicit structure associated with words by the precedence matrix.

**Theorem 3.5.** Let $A$ be an OPA; then an OPG $G$ such that $L(G) = L(A)$ can effectively be built.

**Proof.** Given an OPA $A = \langle \Sigma, M, Q, I, F, \delta \rangle$, we show how to build an equivalent OPG $G$ having OPM $M$. The equivalence between $A$ and $G$ should then be rather obvious.

$G$’s nonterminals are the 4-tuples $(a, q, p, b) \in \Sigma \times Q \times Q \times \Sigma$, written as $(a, q, p, b)$. $G$’s rules are built as follows:

- for every support of type (2.1) of a simple chain, the rule
  \[
  \langle a_0, q_0, a_{n+1} \rangle \rightarrow a_1 a_2 \ldots a_n
  \]
  is in $P$; furthermore, if $a_0 = a_{n+1} = \#$, $q_0$ is initial, and $q_{n+1}$ is final, then $(\# q_0, q_{n+1})$ is in $S$;

- for every support of type (2.2) of a composed chain, add the rule
  \[
  \langle a_0, q_0, a_{n+1} \rangle \rightarrow \Lambda_0 a_1 \Lambda_1 a_2 \ldots a_n \Lambda_n,
  \]
  where, for every $i = 0, 1, \ldots, n$, $\Lambda_i = \langle a_i, q_i, a_i^{\alpha_i+1} \rangle$ if $x_i \neq \varepsilon$ and $\Lambda_i = \varepsilon$ otherwise; furthermore, if $a_0 = a_{n+1} = \#$, $q_0$ is initial, and $q_{n+1}$ is final, then add $(\# q_0, q_{n+1})$ to $S$ and, if $\varepsilon$ is accepted by $A$, add $A \rightarrow \varepsilon$, $A$ being a new axiom not otherwise occurring in any other rule.

Notice that the above construction is effective thanks to the hypothesis of $\overline{\varepsilon}$-acyclicity of the OPM (remember that, as discussed in section 1, this hypothesis could be replaced by weaker ones). This implies that the length of the r.h.s. is bounded (see section 1); on the other hand, the cardinality of the nonterminal alphabet is finite (precisely it is $O(|\Sigma|^2 \cdot |Q|^2)$). Hence there is only a finite number of possible productions for $G$ and only a limited number of chains to be considered.

4. MSO logic characterization. In his seminal paper [12] Büchi provided a logic characterization of regular languages: he defined an MSO syntax on the integers representing the position of characters within a string and, by means of clever arguments, gave algorithms to build a finite-state machine (FSM) recognizing exactly the strings satisfying a given formula and, conversely, to build a formula satisfied by all and only the strings accepted by a given FSM. Subsequently, a rich literature considerably extended his work to more powerful language families—typically, CF [13]—and different logic formalisms, e.g., first-order or tree logics [1, 10, 16]. To the best of our knowledge, MSO logic characterizations of CF languages refer to “visible structure languages,” i.e., to languages whose strings make their syntactic structure immediately visible in their external appearance, such as “tree languages” [39] and VPLs [3] which explicitly refer to this peculiar property in their name. In this section we provide a complete MSO logic characterization of OPLs, which, instead, also include invisible-structure languages, whose syntax trees associated with external strings must be built by means of suitable parsing algorithms, in which the OPM plays a major role.

Similarly to other approaches, in particular to the VPLs one, which in fact are a subclass of OPLs, we begin by defining a suitable binary predicate on the string

\[\text{It is not a coincidence that tree automata [39] have been defined by extending the original finite-state ones.}\]
positions. However the original definition of [4] which states the $\Rightarrow$ relation between the positions of two matching parentheses (calls and returns in VPLs terminology) cannot be naturally extended to the more general case of OPL strings. In fact the $\Rightarrow$ relation between two matching parentheses, which are extremes of the frontier of a subtree, is typically one-to-one (with the exclusion of the particular case of unmatched parentheses which however occur only at the begin and end of a string), whereas, in general, the relation between leftmost and rightmost leaves of an OPL subtree can be many-to-one or one-to-many or both. A further consequence of the more general structure of OPL trees is that, unlike FSMs, tree automata, and VPAs, OPLs are not real-time automata as they may have to perform a series of pop moves without advancing their running head; this in turn produces the effect that, whereas in regular and VPLs each position is associated with a unique state visited by the machine during its behavior, for OPLs the same position may refer to several states—i.e., to several subsets of positions according to Büchi’s approach.

Consequently, the approach we describe here departs from previous ones along two main directions:

- The binary relations between positions referring to a pop operation are attached to the lookback and lookahead positions which in OP parsing embrace the r.h.s. to be reduced; thus, the formal definition of the relation will be based on the notion of chain.

- The sets of positions associated with the different automaton states are subdivided into three, not necessarily disjoint, subsets: one describing the state reached after a push or shift operation, and two to delimit the positions corresponding to each pop operation; in such a way we obtain a unique identification thereof.

4.1. An MSO logic over OP alphabets. Let $(\Sigma, M)$ be an OP alphabet. Let us define a countable infinite set of first-order variables $x, y, \ldots$ and a countable infinite set of MSO (set) variables $X, Y, \ldots$. In the following we adopt the convention to denote first- and second-order variables in boldface italic font.

Definition 4.1 (MSO logic over $(\Sigma, M)$). Let $V_1$ be a set of first-order variables and $V_2$ be a set of second-order (or set) variables. The MSO$_{\Sigma, M}$ (MSO logic over $(\Sigma, M)$) is defined by the following syntax (symbols $\Sigma, M$ will be omitted unless necessary to prevent confusion):

$$\varphi := c(x) \mid x \in X \mid x \leq y \mid x \sim y \mid \neg \varphi \mid \varphi \lor \psi \mid \exists x. \varphi \mid \exists X. \varphi,$$

where$^7$ $c \in \Sigma \cup \{\#\}$, $x, y \in V_1$, and $X \in V_2$.

An MSO formula is interpreted over a $(\Sigma, M)$ string $w$, w.r.t. assignments $\nu_1 : V_1 \rightarrow \{0, 1, \ldots, |w| + 1\}$ and $\nu_2 : V_2 \rightarrow \wp\{0, 1, \ldots, |w| + 1\}$, in the following way.

- $\#w\#, M, \nu_1, \nu_2 \vdash c(x)$ iff $\#w\# = w_1cw_2$ and $|w_1| = \nu_1(x)$.
- $\#w\#, M, \nu_1, \nu_2 \vdash x \in X$ iff $\nu_1(x) \in \nu_2(X)$.
- $\#w\#, M, \nu_1, \nu_2 \vdash x \leq y$ iff $\nu_1(x) \leq \nu_1(y)$.
- $\#w\#, M, \nu_1, \nu_2 \vdash x \sim y$ iff $\#w\# = w_1aw_2bw_3$, $|w_1| = \nu_1(x)$, $|w_1aw_2| = \nu_1(y)$, and $aw_2b$ is a chain $a[w_2]$.
- $\#w\#, M, \nu_1, \nu_2 \vdash \neg \varphi$ iff $\#w\#, M, \nu_1, \nu_2 \not\vdash \varphi$.
- $\#w\#, M, \nu_1, \nu_2 \vdash \varphi_1 \lor \varphi_2$ iff $\#w\#, M, \nu_1, \nu_2 \vdash \varphi_1$ or $\#w\#, M, \nu_1, \nu_2 \vdash \varphi_2$.
- $\#w\#, M, \nu_1, \nu_2 \vdash \exists x. \varphi$ iff $\#w\#, M, \nu_1', \nu_2 \vdash \varphi$, for some $\nu_1'$ with $\nu_1'(y) = \nu_1(y)$ for all $y \in V_1 \setminus \{x\}$.

$^7$This is the usual MSO over strings, augmented with the $\sim$ predicate.
\[ \forall \] such that \[ \phi \] sentence used; then an equivalent OPA is built inductively.

First the formula is rewritten so that no predicate symbols nor first-order variables are only when they are needed (i.e., to give precedence of + over ×).

Note that this property is guaranteed a priori by the structure of the OPM.

To improve readability, we will drop abbreviations in the formulas, such as \[ \# \].

A sentence is a formula without free variables. The language of all strings \( w \in \Sigma^* \) such that \( w \models \phi \) is denoted by \( L(\phi) \):

\[ L(\phi) = \{ w \in \Sigma^* \mid w \models \phi \}. \]

Figure 10 illustrates the meaning of the \( \bowtie \) relation with reference to the string of Figure 1: we have \( 0 \bowtie 2, 2 \bowtie 4, 5 \bowtie 7, 7 \bowtie 9, 5 \bowtie 9, 4 \bowtie 10, 2 \bowtie 10, \) and \( 0 \bowtie |w| + 1. \)

Example 7. Consider the language of Example 1. The following sentence states that all parentheses are well-matched:

\[ \forall x \forall y \left( x \bowtie y \Rightarrow ((x + 1) \Rightarrow (y - 1) \wedge \neg \exists z \left( z < y \wedge x \bowtie z \wedge \neg \exists v \left( x < v \wedge v \bowtie y \right) \right)) \right). \]

Note that this property is guaranteed a priori by the structure of the OPM.

The following sentence instead defines the language where parentheses are used only when they are needed (i.e., to give precedence of + over ×):

\[ \forall x \forall y \left( \left( x \bowtie y \wedge ((x + 1) \wedge (y - 1)) \Rightarrow (x \bowtie y) \wedge \exists z \left( x + 1 < z < y - 1 \wedge (z \wedge \neg \exists u \exists v \left( u < v \wedge y - 1 \wedge (u \wedge v) \right) \right)) \right) \right). \]

The following theorem states the main result of this section.

Theorem 4.2. A language \( L \) over \( (\Sigma, M) \) is an OPL iff there exists an MSO sentence \( \phi \) such that \( L = L(\phi) \).

The proof is constructive and structured in the following two subsections.

4.2. From MSO to OPAs.

Statement 4.1. Let \( (\Sigma, M) \) be an OP alphabet and \( \varphi \) be an MSO sentence. Then \( L(\varphi) \) can be recognized by an OPA over \( (\Sigma, M) \).

Proof. The proof follows the one by Thomas [40] and is composed of two steps: first the formula is rewritten so that no predicate symbols nor first-order variables are used; then an equivalent OPA is built inductively.
Let Σ be \{a_1, a_2, \ldots, a_n\}. For each predicate symbol \(a_i\) we introduce a fresh set variable \(X_i\), therefore formula \(a_i(x)\) will be translated into \(x \in X_i\). Following the standard construction of \([40]\), we also translate every first-order variable into a fresh second-order variable with the additional constraint that the set it represents contains exactly one position. The only difference is that formulas like \(x \sim y\) will be translated into formulas \(X_i \sim X_j\), where \(X_i, X_j\) are singleton sets. In this case, the semantics of \(\sim\) is naturally extended to second-order variables that are singletons.

Let \(\varphi'\) be the formula obtained from \(\varphi\) by such a translation, and consider any subformula \(\psi\) of \(\varphi'\): let \(X_1, X_2, \ldots, X_n, X_{n+1}, \ldots, X_{n+m(\psi)}\) be the (second-order) free variables appearing in \(\psi\). Recall that \(X_1, \ldots, X_n\) represent symbols in Σ, hence they are never quantified.

As usual we interpret formulas over strings; in this case we use the alphabet

\[\Lambda(\psi) = \left\{ \alpha \in \{0, 1\}^{n+m(\psi)} \mid \exists i \text{ s.t. } 1 \leq i \leq n, \alpha_i = 1 \right\}.\]

A string \(w \in \Lambda(\psi)^*\), with \(|w| = \ell\), is used to interpret \(\psi\) in the following way: the projection over the \(j\)th component of \(\Lambda(\psi)\) gives a valuation \(\{1, 2, \ldots, \ell\} \to \{0, 1\}\) of \(X_j\) for every \(1 \leq j \leq n + m(\psi)\).

For any \(\alpha \in \Lambda(\psi)\), the projection of \(\alpha\) over the first \(n\) components encodes a symbol in Σ, denoted as \(\text{symbol}(\alpha)\). The matrix \(M\) over Σ can be naturally extended to the OPM \(M(\psi)\) over \(\Lambda(\psi)\) by defining \(M(\psi)_{\alpha, \beta} = M_{\text{symbol}(\alpha), \text{symbol}(\beta)}\) for any \(\alpha, \beta \in \Lambda(\psi)\).

We now build an OPA \(A\) equivalent to \(\varphi'\). The construction is inductive on the structure of the formula: first we define the OPA for all atomic formulas. We give here only the construction for \(\sim\), since for the other ones the construction is standard and is the same as in \([40]\).

Figure 11 represents the OPA for atomic formula \(\psi = X_i \sim X_j\) (notice that \(i, j > n\), and that both \(X_i\) and \(X_j\) are singleton sets). For the sake of brevity, we use notation \([X_i]\) to represent the set of all tuples \(\Lambda(\psi)\) having the \(i\)th component equal to 1; notation \([X]\) represents the set of all tuples in \(\Lambda(\psi)\) having both \(i\)th and \(j\)th components equal to 0.

The semantics of \(\sim\) requires for \(X_i \sim X_j\) that there must be a chain \(^a[w_2]\) in the input word, where \(a\) is the symbol at the only position in \(X_i\), and \(b\) is the symbol at the only position in \(X_j\). By definition of chain, this means that \(a\) must be read, hence in the position represented by \(X_i\) the automaton performs either a push or a...
shift move (see Figure 11, from state $q_0$ to $q_1$), as pop moves do not consume input. After that, the automaton must read $w_2$. In order to process the chain $a[w_2]^b$, reading $w_2$ must start with a push move (from state $q_1$ to state $q_2$), and it must end with one or more pop moves, before reading $b$ (i.e., the only position in $X_j$—going from state $q_3$ to $q_F$).

This means that the automaton, after a generic sequence of moves corresponding to visiting an irrelevant (for $X_i \setminus X_j$) portion of the syntax tree, when reading the symbol at position $X_i$ performs either a push or a shift move, depending on whether $X_i$ is the position of a leftmost leaf of the tree or not. Then it visits the subsequent subtree ending with a pop labeled $q_1$; at this point, if it reads the symbol at position $X_j$, it accepts anything else that follows the examined fragment.

Then, a natural inductive path leads to the construction of the automaton associated with a generic MSO logic formula: the disjunction of two subformulas can be obtained by building the union automaton of the two corresponding automata; similarly for negation. The existential quantification of $X_i$ is obtained by projection erasing the $i$th component; since OPAs are closed under alphabetical homomorphisms preserving the OPM (see Statement 1.2), and since the OPM is determined only by the first $n$ components of the alphabet’s elements which are never erased by quantification, such a projection produces a well-defined automaton for any $\psi$. Finally, the alphabet of the automaton equivalent to $\varphi'$ is $\Lambda(\varphi') = \{0, 1\}^n$, which is in bijection with $\Sigma$. □

4.3. From OPAs to MSO.

When considering a chain $a[w]^b$ we assume $w = w_0a_1w_1 \ldots a_tw_t$, with $a[a_1a_2\ldots a_t]^b$ being a simple chain (any $w_2$ may be empty). We denote by $s_g$ the position of symbol $a_g$ for $g = 1, 2, \ldots, t$ and set $a_0 = a$, $s_0 = 0$, $a_{t+1} = b$, and $s_{t+1} = |w| + 1$. Furthermore, we define the following shortcut notations:

$$x \circ y := \bigvee_{M_{a,h}=0} a(x) \land b(y) \text{ for } \circ \in \{<, \div, \rangle\}.$$  

$$\text{Tree}(x, z, v, y) := x \land \left( (x + 1 = z \lor x \lor z) \land \lnot \exists t(z < t < y \land x \lor y) \right) \land \left( (v + 1 = y \lor v \lor y) \land \lnot \exists t(x < t < v \lor t \lor y) \right).$$

If $x \lor y$ then there exist (unique) $z$ and $v$ such that $\text{Tree}(x, z, v, y)$ is satisfied. In particular, if $w$ is the body of a simple chain, then $0 \lor w + 1$ and $\text{Tree}(0, 1, t, t + 1)$ are satisfied; if it is the body of a composed chain, then $0 \lor w + 1$ and $\text{Tree}(0, s_1, s_t, s_{t+1})$ are satisfied. If $w_0 = \varepsilon$ then $s_1 = 1$, and if $w_t = \varepsilon$ then $s_t = |w|$. In the example of Figure 10 relations $\text{Tree}(2, 3, 3, 4)$, $\text{Tree}(2, 4, 4, 10)$, $\text{Tree}(4, 5, 9, 10)$, $\text{Tree}(5, 7, 7, 9)$ are satisfied, among others.

Statement 4.2. Let $(\Sigma, M)$ be an OP alphabet and $A$ be an OPA over $(\Sigma, M)$. Then there exists an MSO sentence $\varphi$ such that $L(A) = L(\varphi)$.

Proof. Let $A = (\Sigma, M, Q, q_0, F, \delta)$ be deterministic (this simplifying assumption does not cause loss of generality, since nondeterministic OPAs are equivalent to deterministic ones by Theorem 2.4). Without loss of generality we also assume that the transition function of $A$ is total. We build an MSO sentence $\varphi$ such that $L(A) = L(\varphi)$. The main idea for encoding the behavior of the OPA is based on assigning the states visited during its run to positions along the same lines stated by Büchi in Thomas [40] and extended for VPLs [4]. Unlike finite-state automata and VPAs, however, OPAs do not work on-line. Hence, it is not possible to assign a single state to every position. Let $Q = \{q_0, q_1, \ldots, q_N\}$ be the states of $A$ with $q_0$ initial; as usual, we will use...
second-order variables to encode them. We shall need three different sets of second-order variables, namely, \( A_0, A_1, \ldots, A_N, B_0, B_1, \ldots, B_N, \) and \( C_0, C_1, \ldots, C_N \). Set \( A_i \) contains those positions of word \( w \) where state \( q_i \) may be assumed after a shift or push transition, i.e., after a transition that “consumes” an input symbol. Sets \( B_i \) and \( C_i \) encode a pop transition concluding the reading of the body \( w_0q_1w_1\ldots q_kw_l \) of a chain whose support ends in a state \( q_l \); set \( B_i \) contains the position of symbol \( a \) that precedes the corresponding push, whereas \( C_i \) contains the position of \( a_l \), which is the symbol on top of the stack when the automaton performs the pop move. Figure 12 presents such sets for the example automaton of Figure 5, with the same input as in Figure 10. Notice that each position, except the last one, belongs to exactly one \( A_i \), whereas it may belong to several \( B_i \) and at most one \( C_i \).

Then, sentence \( \varphi \) is defined as follows:

\[
\varphi := \exists e \ (\exists A_0, A_1, \ldots, A_N \\
\exists B_0, B_1, \ldots, B_N \\
\exists C_0, C_1, \ldots, C_N) \left( \text{Start}_0 \wedge \varphi_{\text{init}} \wedge \bigvee_{q_f \in F} \text{End}_f \right),
\]

where the first and last subformulas encode the initial and final states of the run, respectively; formula \( \varphi_{\text{init}} \) is defined as \( \varphi_{\text{push}} \wedge \varphi_{\text{shift}} \wedge \varphi_{\text{pop}} \) and encodes the three transition functions of the automaton, which are expressed as the conjunction of forward and backward formulas. Variable \( e \) is used to refer to the end of a string.

To complete the definition of \( \varphi \), we incrementally introduce more notation:

\[
\text{Succ}_k(x, y) := x + 1 = y \wedge x \in A_k,
\]

\[
\text{Next}_k(x, y) := x \wedge y \wedge x \in B_k \wedge \exists z, v \ (\text{Tree}(x, z, v, y) \wedge v \in C_k),
\]

\[
Q_i(x, y) := \text{Succ}_i(x, y) \lor \text{Next}_i(x, y).
\]

The shortcut \( Q_i(x, y) \) is used to represent that \( A \) is in state \( q_i \) when at position \( x \) and the next position to read, possibly after scanning a chain, is \( y \). Since the automaton is not real time, we must distinguish between push and shift moves (case \( \text{Succ}_i(x, y) \)),

\[
\text{Start}_0 := \forall v \ (v = \text{null} \lor \exists x, y \ (x = y \wedge Q_i(x, y) \wedge v = \text{null})).
\]
and pop moves (case Next$_i(x, y)$). For instance, with reference to Figures 10 and 12, Succ$_2(5, 6)$, Next$_3(5, 9)$, and Next$_3(5, 7)$ hold.

The shortcuts representing the initial and final states of the parsing of a string of length $e$ are defined as follows:

\[
\text{Start}_i := 0 \in A_i \land -\bigvee_{j \neq i} (0 \in A_j),
\]

\[
\text{End}_f := -\exists y(e + 1 < y) \land \text{Next}_f(0, e + 1) \land -\bigvee_{j \neq f} (\text{Next}_f(0, e + 1)).
\]

$\varphi_{\delta_{\text{push}}}$ is the conjunction of the following two formulas. The former states the sufficient condition for a position to be in a set $A_i$, when performing a push move:

\[
\varphi_{\text{push}_{i,w}} := \forall x, y \prod_{i=0}^{N} \prod_{k=0}^{N} (x \preceq y \land c(y) \land Q_i(x, y) \land \delta_{\text{push}}(q_i, c) = q_k \Rightarrow y \in A_k).
\]

The latter formula states the symmetric necessary condition:

\[
\varphi_{\text{push}_{i,bw}} := \forall x, y \prod_{i=0}^{N} \prod_{k=0}^{N} (x \preceq y \land c(y) \land y \in A_k \land (x + 1 = y \lor x \cap y) \Rightarrow \bigvee_{i=0}^{N} (Q_i(x, y) \land \delta_{\text{push}}(q_i, c) = q_k)).
\]

$\varphi_{\delta_{\text{shift}}}$ is defined analogously, w.r.t. shift moves instead of push moves:

\[
\varphi_{\text{shift}_{i,w}} := \forall x, y \prod_{i=0}^{N} \prod_{k=0}^{N} (x \preceq y \land c(y) \land \delta_{\text{shift}}(q_i, c) = q_k \Rightarrow y \in A_k),
\]

\[
\varphi_{\text{shift}_{i,bw}} := \forall x, y \prod_{i=0}^{N} \prod_{k=0}^{N} (x \preceq y \land c(y) \land y \in A_k \land (x + 1 = y \lor x \cap y) \Rightarrow \bigvee_{i=0}^{N} (Q_i(x, y) \land \delta_{\text{shift}}(q_i, c) = q_k)).
\]

Finally, to define $\varphi_{\delta_{\text{pop}}}$ we introduce the shortcut Tree$_{i,j}(x, z, v, y)$, which represents the fact that $A$ is ready to perform a pop transition from state $q_i$ having on top of the stack state $q_j$; such pop transition corresponds to the reduction of the portion of string between positions $x$ and $y$ (excluded):

\[
\text{Tree}_{i,j}(x, z, v, y) := \text{Tree}(x, z, v, y) \land Q_i(v, y) \land Q_j(x, z).
\]

Formula $\varphi_{\delta_{\text{pop}}}$ is thus defined as the conjunction of three formulas. As before, the forward formula gives the sufficient conditions for two positions to be in the sets $B_k$ and $C_k$, when performing a pop move, and the backward formulas state symmetric necessary conditions:

\[
\varphi_{\text{pop}_{i,w}} := \forall x, z, v \prod_{i=0}^{N} \prod_{j=0}^{N} \prod_{k=0}^{N} (\text{Tree}_{i,j}(x, z, v, y) \land \delta_{\text{pop}}(q_i, q_j) = q_k \Rightarrow x \in B_k \land v \in C_k),
\]

\[
\varphi_{\text{pop}_{i,bw,B}} := \forall x \prod_{k=0}^{N} (x \in B_k \Rightarrow \exists y, z, v \prod_{i=0}^{N} \prod_{j=0}^{N} \text{Tree}_{i,j}(x, z, v, y) \land \delta_{\text{pop}}(q_i, q_j) = q_k),
\]

\[
\varphi_{\text{pop}_{i,bw,C}} := \forall v \prod_{k=0}^{N} (v \in C_k \Rightarrow \exists x, y, z \prod_{i=0}^{N} \prod_{j=0}^{N} \text{Tree}_{i,j}(x, z, v, y) \land \delta_{\text{pop}}(q_i, q_j) = q_k).
\]
Now notice that \( \varphi \equiv \bigvee_{q_f \in F} \psi_{0,f} \), where

\[
\psi_{t,k} := \exists \psi B_0, B_1, \ldots, B_N \quad \psi \varphi_0 \wedge \varphi(B_0, B_1, \ldots, B_N) \quad (\text{Start}_i \wedge \varphi_\delta \wedge \text{End}_k),
\]

\[
\exists C_0, C_1, \ldots, C_N.
\]

Hence, the proof that \( L(A) = L(\varphi) \) is a direct consequence of the following Lemmas 4.3 and 4.4, stating that \( w \models \psi_{t,k} \) if \( q_i \xrightarrow{w} q_k \) in \( A \), for every word \( w \) compatible with \((\Sigma, M)\).

**Lemma 4.3.** Let \( w \) be the body of a chain \( w \). If \( q_i \xrightarrow{w} q_k \) in \( A \), then \( w \models \psi_{t,k} \).

**Proof.** We prove the lemma by induction on the depth of chains. Note that, even if \( A \) is deterministic, some chains could have different supports. However, we will show that every support produces exactly one assignment that satisfies \( \psi_{t,k} \).

Let \( w \) be the body of a simple chain with support

\[
q_i = q_{t_0} \xrightarrow{a_1} q_{t_1} \xrightarrow{a_2} \cdots \xrightarrow{a_{\ell}} q_{t_{\ell}} \xrightarrow{q_{f_0}} q_k.
\]

We prove that \( w \models \psi_{t,k} \) for \( e, A_0, A_1, \ldots, A_N, B_0, \ldots, B_N, C_0, \ldots, C_N \) defined as follows. First-order variable \( e \) equals \( |w| \), \( B_h \) is empty except for \( B_h = \{0\} \), \( C_h \) is empty except for \( C_k = \{\ell\} \); for every \( 0 \leq x \leq \ell \), let \( A_h \) contain \( i \) if \( t_x = h \) (i.e., \( x \in A_{t_x} \)), and this also implies \( Q_{i_h}(x, x + 1) \). Then \( \text{Start}_i \) and \( \text{End}_k \) are satisfied trivially since \( \text{Tree}(0, 1, \ell, \ell + 1) \) holds. We now prove that also \( \varphi_{\delta_{\text{push}}}, \varphi_{\delta_{\text{shift}}}, \) and \( \varphi_{\delta_{\text{pop}}} \) are satisfied; we omit considering all cases where the antecedents are false.

- \( \varphi_{\text{push}} \) is satisfied for \( x = 0 \) and \( y = 1 \) since we have \( \# < a_1, Q_{0(0, 1)}, 1 \in A_{t_1}, \) and \( \delta_{\text{push}}(q_{t_1}, a_1) = q_{t_1} \).
- \( \varphi_{\text{shift}} \) is satisfied \( \forall 1 \leq x < \ell \) and \( y = x + 1 \) since we have \( a_y(y), a_x = a_y, Q_{1_h}(x, y), y \in A_{t_x} \), and \( \delta_{\text{shift}}(q_{t_x}, a_y) = q_{t_y} \).
- \( \varphi_{\text{pop}} \) is satisfied for \( x = 0 \) and \( y = |w| + 1 = \ell + 1 \) since we have \( \text{Tree}_{t_x,i}(0, 1, \ell, \ell + 1), 0 \in B_k, \ell \in C_k, \) and \( \delta_{\text{pop}}(q_{t_k}, q_i) = q_k \).

Let now \( w \) be the body of a composed chain with support

\[
q_i = q_{t_0} \xrightarrow{w_0} q_{f_0} \xrightarrow{a_1} q_{t_1} \xrightarrow{w_1} q_{f_1} \xrightarrow{a_2} \cdots \xrightarrow{w_\ell} q_{f_\ell} \xrightarrow{\#} q_k.
\]

We prove that \( w \models \psi_{t,k} \) for a suitable assignment. By the inductive hypothesis, for every \( g = 0, 1, \ldots, \ell \) such that \( \omega_g \neq \varepsilon \) we have \( w_g \models \psi_{t_g,f_g} \). Let \( A_{0g}, \ldots, A_{Ng}, B_{0g}, \ldots, B_{Ng}, C_{0g}, \ldots, C_{Ng} \) be the (naturally shifted versions of) an assignment that satisfies \( \psi_{t_g,f_g} \). In particular this implies \( s_g \in A_{t_g} \), Next\(_{f_g}(s_g, s_{g+1}) \), and \( s_g \in A_{t_g} \cup B_{f_g} \) for each \( g \) such that \( \omega_g \neq \varepsilon \). Then define \( A_h, B_h, C_h \) as follows. Let \( A_h \) include all \( A_{0h}, B_{0h}, C_{0h} \) include all \( B_{0h}, C_{0h} \) include all \( C_{0h} \). Also let \( B_h \) contain \( s_0, C_h \) contain \( s_1, A_h \) contain \( s_g \) whenever \( w_g \) is empty; in particular, this implies \( Q_{1h}(s_g, s_{g+1}) \) for every \( 0 \leq g < \ell \). Finally, \( e \) is defined as the length of \( w \).

Then we show that \( \psi_{t_g} \) is satisfied by checking every subformula. Start, End, and \( \text{End}_k \) are satisfied trivially since \( \text{Tree}(0, s_1, s_{s_1} + 1) \) holds. By the inductive hypothesis, all other axioms are satisfied within every \( w_g \). Thus, we only have to prove that they are satisfied in positions \( s_g \) for \( 0 \leq g < \ell \). We omit considering all cases where the antecedents are false.

- \( \varphi_{\text{push}} \) is satisfied for \( x = 0 \) and \( y = s_1 \) since we have \( a_1(s_1), \# < a_1 Q_{f_0}(0, s_1), s_1 \in A_{t_1}, \) and \( \delta_{\text{push}}(q_{f_0}, a_1) = q_{t_1} \).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
- \( \varphi_{\text{shift}} \) is satisfied for all \( x = s_g \) and \( y = s_{g+1} \) with \( 1 \leq g < \ell \) since we have 
\[ a_g(s_g), a_g = a_{g+1}, Q_f(s_g, s_{g+1}), s_g \in A_g, \text{ and } \delta_{\text{shift}}(q_f, a_g) = q_g. \]
- \( \varphi_{\text{pop}} \) is satisfied for \( x = 0 \) and \( y = |w| + 1 \) since we have \( \text{Tree}_{f_x, f_0}(0, s_1, s_\ell, |w| + 1), 0 \in B_k, \ell \in C_k, \text{ and } \delta_{\text{pop}}(q_k, q_1) = q_k. \)

Hence \( w \models \psi_{i,k} \) for every \( w \) with a suitable support, and this concludes the proof. \( \square \)

**Lemma 4.4.** Let \( w \) be the body of a chain \( \# |w| \). If \( w \models \psi_{i,k} \) then \( q_i \overset{w}{\leadsto} q_k \) in \( A \).

**Proof.** Let \( e = |w|, A_0, \ldots, A_N, B_0, \ldots, B_N, C_0, \ldots, C_N \) be an assignment that satisfies \( \psi_{i,k} \). In particular, this implies \( 0 \in A_1 \land \text{Next}_k(0, |w| + 1) \), and such \( i, k \) are unique by definition of \( \text{Start}_i \) and \( \text{End}_k \). Then the following properties hold.

1. For each \( 0 \leq x \leq |w| \), there exists a unique index \( i \) such that \( \text{Succ}_i(x, x + 1) \) holds true. This can be proved by induction on \( x \) by applying the formulas for \( \delta_{\text{push}} \) and \( \delta_{\text{shift}} \).
2. For each \( x, y \) such that \( x \land y \), let \( z, v \) such that \( \text{Tree}(x, z, v, y) \) holds, then there exists a unique pair of indices \( i, j \) such that \( \text{Tree}_{i,j}(x, z, v, y) \) holds, and there exists a unique index \( k \) such that \( \text{Next}_k(x, y) \). This can be proved by induction on the depth of the chain between positions \( x \) and \( y \), by applying the formulas for \( \delta_{\text{pop}} \) and property 1.

Moreover, if \( \text{Tree}_{i,j}(x, z, v, y) \) holds, then \( \text{Next}_k(x, y) \) holds iff \( \delta_{\text{pop}}(q_i, q_j) = q_k. \)

Hence, by properties 1 and 2, for each \( x, y \) such that \( x + 1 = y \) or \( x \land y \), there exists a unique index \( i \) such that \( Q_i(x, y) \) holds true.

Now, for every \( g \) let \( t_g \) be the index such that \( g \in A_{t_g}, t_g \) is unique by property 1 and in particular \( t_0 = i \).

We proceed by induction on the depth \( h \) of \( w \). Let \( h = 1 \) and \( w = a_1a_2 \ldots a_\ell \) be the body of a simple chain. In this case \( t_g \) is the unique index such that \( \text{Succ}_{t_g}(g, g + 1) \). Then, by \( \varphi_{\text{push}, f_w} \) with \( y = 1 \), we have \( \delta(q_{t_0}, a_1) = q_{t_1} \); and by \( \varphi_{\text{shift}, f_w} \) with \( 1 \leq \ell < k \), we have \( \delta_{\text{shift}}(q_{t_0}, a_{g+1}) = q_{t_{g+1}} \). Moreover, since \( \text{Tree}_{t_{0}, t_0}(0, 1, \ell, \ell + 1) \land \text{Next}_k(0, \ell, \ell + 1), \) we get \( \delta(q_{t_0}, q_{t_0}) = q_k \) by property 2. Hence we have built a support of the type (4.2).

Let now \( h > 1 \) and \( w = w_0a_1w_1 \ldots a_\ell w_\ell \). For \( 0 \leq g \leq \ell \), since \( s_g \land s_{g+1} \cup s_{g+1} = s_{g+1} \), by properties 1 and 2 above there exists a unique index \( f_g \) such that \( Q_{f_g}(s_g, s_{g+1}) \) holds. Notice that \( w_g = \varepsilon \) implies \( f_g = t_g \), otherwise we have \( w_g \models \varphi_{t_g, f_g} \) and, by the inductive hypothesis, there exists a support \( q_{t_0} \overset{s_g}{\leadsto} q_f \) in \( A \).

Thus, for every \( 0 \leq g < \ell \), by applying \( \varphi_{\text{push}, f_w} \) with \( y = s_{g+1} \) we get \( \delta(q_f, a_{g+1}) = q_{t_{g+1}} \). Moreover, since \( \text{Tree}_{f_x, f_0}(0, s_1, s_\ell, |w| + 1) \land \text{Next}_k(0, |w| + 1), \) by property 2 above we get \( \delta(q_{t_0}, q_{t_0}) = q_k \). Hence we have built a support of type (4.3) and this concludes the proof. \( \square \)

II: Operator precedence \( \omega \)-languages.

Languages of infinite-length strings, called \( \omega \)-languages, have been introduced to model nonterminating processes; thus they are becoming more and more relevant nowadays when most applications are “ever-running”, often in a distributed environment. Again, the foundations of the theory of \( \omega \)-languages are due to the pioneering work by Büchi [12] and others [32, 30, 37, 9]. Büchi, in particular, investigated their main algebraic properties in the context of finite-state machines, pointing out commonalities and differences w.r.t. the finite-length counterpart [12, 40]. His work has
then been extended to larger classes of languages; among them, again noticeably, the
class of VPLs; and again, in this part we face the same job for the class of OPLs.
OPLs, in fact, are not only useful to model programs, which are typically of finite
length, but are also well-suited to formalize possibly never-ending sequences of events:
for instance, the previous Example 5 can be naturally extended to model the behavior
of a database that is never put off.

This part is organized as follows. In section 5 we first extend to \( \omega \)-languages
a few basic definitions given in Part I for finite-length languages and generalize to
OPAs the classical accepting criteria for \( \omega \)-languages, then we show by means of
an example the usefulness of \( \omega \)-OPAs to model and analyze various system types;
section 6 shows the relations between the various classes of \( \omega \)-OPLs classified according
to the acceptance criteria defined in the previous section; section 7 shows which
closure properties are preserved and which ones are lost when moving from finite-
length languages to the various classes of \( \omega \)-languages; finally, section 8 extends to
\( \omega \)-languages the characterization in terms of MSO logic.

5. Basic definitions of \( \omega \)-languages. Preliminarily we introduce some further
properties related to chains that are necessary when chains occur within infinite words.

Definition 5.1. Let \((\Sigma, M)\) be a precedence alphabet and \(w\) a word on \(\Sigma\) com-
patible with \(M\):

- A chain in \(w\) is maximal if it does not belong to a larger composed chain. In
  a finite word \(w\) preceded and ended by \#, only the outmost chain \#\(|w|\#\) is
  maximal.
- An open chain is a sequence of symbols \(b_0 < a_1 \leq a_2 \leq \cdots \leq a_n\) for \(n \geq 1\).
- A letter \(a \in \Sigma\) in a word \#\(w\) with \(w \in \Sigma^*\) compatible with \(M\), is pending if
  it does not belong to the body of a chain. In a word \(w\) preceded and ended by
  \#, there are no pending letters.

Furthermore, we generalize in a natural way to the infinite case the notion of a
string compatible with an OPM: given a precedence alphabet \((\Sigma, M)\), we say that
an \(\omega\)-word \(w\) is compatible with the OPM \(M\) if every prefix of \(w\) is compatible with
\(M\). We denote by \(L_M \subseteq \Sigma^\omega\) the \(\omega\)-language comprising all infinite words \(x \in \Sigma^\omega\)
compatible with \(M\).

Next, we adopt for OPAs operating on infinite strings the same acceptance criteria
that have been adopted in the literature for regular and other classes of languages.

Definition 5.2 (Büchi OP \(\omega\)-automaton). A nondeterministic Büchi OP \(\omega\)
automaton \((\omega\text{OPA})\) is given by a tuple \(A = (\Sigma, M, Q, I, F, \delta)\), where \(\Sigma, Q, I, F, \delta\)
are defined as for OPAs; the OPM \(M\) is restricted to be a \(|\Sigma \cup \{\#\}| \times |\Sigma|\) array, since
\(\omega\)-words are not terminated by the delimiter \#.

Configurations and (infinite) runs are defined as for OPAs on finite-length words.

Then, let \(\mathbb{P}^i\) be a shorthand for “there exist infinitely many \(i\)'s” and let \(\rho\) be a run
of the automaton on a given word \(x \in \Sigma^\omega\). Define \(\text{Inf}(\rho) = \{q \in Q \mid \exists \omega^i \langle \beta_i, q_i, x_i \rangle \in \rho, q_i = q\}\) as the set of states that occur infinitely often in configurations in \(\rho\). A run
\(\rho\) of an \(\omega\)-OPA on an infinite word \(x \in \Sigma^\omega\) is successful iff there exists a state \(q_f \in F\)
such that \(q_f \in \text{Inf}(\rho). A\) accepts \(x \in \Sigma^\omega\) iff there is a successful run of \(A\) on \(x\). The
\(\omega\)-language recognized by \(A\) is \(L(A) = \{x \in \Sigma^\omega \mid A\) accepts \(x\}\).

The classical notion of acceptance for Muller automata can be likewise defined
for OPAs.

Definition 5.3 (Muller OP \(\omega\)-automaton). A nondeterministic Muller OPA
\((\omega\text{OPMA})\) is a tuple \((\Sigma, M, Q, I, T, \delta)\), where \(\Sigma, M, Q, I, \delta\) are defined as for \(\omega\text{OPBAs}\)
and \(T\) is a collection of subsets of \(Q, T \subseteq \mathcal{P}(Q)\), called the table of the automaton.
A run $\rho$ of an $\omega$OPMA on an infinite word $x \in \Sigma^\omega$ is successful if $\text{Inf}(\rho) \in T$, i.e., the set of states occurring infinitely often in the configurations of $\rho$ is a set in the table $T$.

Definition 5.4. A nondeterministic Büchi OPA accepting with empty stack ($\omega$OPBEA) is a variant of $\omega$OPBA where a run $\rho$ is successful if there exists a state $q_f \in T$ such that configurations with stack $\bot$ and state $q_f$ occur infinitely often in $\rho$.

Thus, a run of an $\omega$OPBEA is successful if the automaton traverses final states with an empty stack infinitely often. We will use the following simple normal form for $\omega$OPBEA.

Definition 5.5. An $\omega$OPBEA is in normal form if the set of states is partitioned into two types of states that are always visited with an empty stack and states that are never visited with an empty stack.

For all the above classes of automata, say, $\omega$-XXX, their deterministic counterpart $\omega$-DXXX is defined as usual.

Example 8 (managing interrupts). Consider a software system that is designed to work forever and must serve requests issued by different users but subject to interrupts. Precisely, assume that the system manages two types of “normal operations” $a$ and $b$, and two types of interrupts with different levels of priority.

We model its behavior by introducing an alphabet with two pairs of calls and returns, $\text{call}_a$, $\text{call}_b$, $\text{ret}_a$, $\text{ret}_b$, for operations $a$ and $b$ and symbols $\text{int}_1$, $\text{serve}_1$ denoting the lower level interrupt and its serving, respectively, and $\text{int}_2$, $\text{serve}_2$ denoting the higher level ones. Not only do both interrupts discard possible pending calls not already matched by corresponding returns, but also the serving of a higher priority interrupt erases possible pending requests for lower priority ones, but not those that occurred before the higher priority interrupt just served: thus, a sequence such as $\text{int}_1\text{int}_2\text{int}_1\text{int}_1\text{serve}_2$ should produce popping the second and third $\text{int}_1$ without matching them, to match immediately $\text{int}_2$ with $\text{serve}_2$, but would leave the first occurrence of $\text{int}_1$ still pending; the next $\text{serve}_1$, if any, would match it, whereas possible further $\text{serve}_1$ would remain unmatched. Furthermore neither calls to, nor returns from, operations $a$ and $b$ can occur while any interrupt is pending.

Table 3 shows an OPM that assigns to sequences on the above alphabet a structure compatible with the described priorities. Then, a suitable $\omega$-automaton can specify further constraints on such sequences; for instance the $\omega$OPBA of Figure 13 restricts the set of $\omega$-sequences compatible with the matrix by imposing that all $\text{int}_2$ are eventually served by a corresponding $\text{serve}_2$; furthermore lower priority interrupts are not just discarded when a higher priority one is pending but they are simply disabled, i.e., they are not accepted as a correct system behavior.
For instance, the ω-word \(call_a\ int_1\ int_2\ int_1\ldots\) is not accepted by the ωOPBA because \(int_1\) is not accepted from state \(q_2\) reached after reading \(int_2\); similarly, \(call_a\ int_1\ int_2\ serve_2\ call_a\) is rejected since, after serving \(int_2\) the automaton would be back in state \(q_1\) with \(int_1\) pending (the prefix \(call_a\ int_1\ int_2\ serve_2\) is compatible with the OPM and \(int_1\) is pending therein) but no \(call_a\) is admitted in \(q_1\) since there is no precedence relation between \(int_1\) and \(call_a\). On the contrary the ω-word \(call_a\ int_1\(\int_2\ serve_2\serve_1\ call_a\ call_a\ ret_a\)^ω\) is accepted: in fact the automaton reaches \(q_1\) after reading \(call_a\) (and popping it) followed by \(int_1\); then, after receiving and serving the higher priority interrupt, it would serve the pending instance of \(int_1\) returning to \(q_0\); from this point on it would enter an infinite loop during which it would process the input string \((call_a\ call_a\ ret_a\ int_2\ serve_2\ serve_1)^\omega\) traversing the states \(q_0\ call_a\ q_0\ call_a\ q_0\ ret_a\ q_0\ int_1\ q_2\ serve_2\ q_2\ \rightarrow q_0\ \rightarrow q_0\ serve_2\) \(q_1\Rightarrow q_0\Rightarrow q_0\) leaving the first \(call_a\) and \(serve_1\) unmatched. Notice that all finite prefixes \(call_a\ int_1\(\int_2\ serve_2\ serve_1\ call_a\ call_a\ ret_a\)^n\ int_2\ serve_2\ serve_1\ call_a\ call_a\), with \(n > 0\), end with the open chain \(call_a\ < call_a\). Finally, observe that the automaton would accept some strings beginning with \(serve_1\) which might appear somewhat counterintuitive but is consistent with the general philosophy of admitting unmatched elements; it would be easy, however, to forbid such a string beginning.

We call \(L\text{interrupt}\) the language recognized by this ωOPBA.

A more sophisticated policy that could easily be formalized by means of a suitable ω-automaton is a “weak fairness requirement” imposing that, after a first \(call_a\) not matched by \(ret_a\) but interrupted by a \(int_1\) or \(int_2\), a second \(call_a\) cannot be interrupted by a new lower priority interrupt \(int_1\) (but can still be interrupted at any time by higher priority ones).

This example too retains some typical features of VPLs, namely, the possibility of having unmatched calls or returns but, again, it strongly generalizes them in that unmatched elements can occur in various places of the whole string, e.g., due to the occurrence of interrupts or other exceptional events.

Further examples illustrating the modeling capabilities of OPLs both on finite and infinite strings are reported in [35].

6. Relationships among classes of ωOPLs. Here we study the relationships among languages recognized by the different classes of OP ω-automata and visibly pushdown ω-automata (with Büchi acceptance criterion), denoted as ωBVPA. Such
relations are summarized by the diagram in Figure 14, where solid lines denote strict inclusion and dashed lines link classes that are not comparable.

In the following, we first present the proofs of the weak containment relations holding among the various classes: most of them follow trivially from the definitions, except for the equality between \( \mathcal{L}(\omega \text{OPBA}) \) and \( \mathcal{L}(\omega \text{OPMA}) \). Then we will prove strict inclusions and incomparability relations by means of a suitable set of examples that separate the various classes.

6.1. Weak inclusion results.

**Theorem 6.1.** The following inclusion relations hold:

\[
\mathcal{L}(\omega \text{BVPA}) \subseteq \mathcal{L}(\omega \text{OPBA}), \quad \mathcal{L}(\omega \text{DBVPA}) \subseteq \mathcal{L}(\omega \text{DOPBA}).
\]

**Proof.** Let \( A = (Q_A, I_A, \Gamma_A, \delta_A, F_A) \) be an \( \omega \text{BVPA}^8 \) over a partitioned alphabet \( \Sigma = (\Sigma_c, \Sigma_r, \Sigma_i) \). An \( \omega \text{OPBA} \) \( B \) that recognizes the same language as \( A \) is defined in a straightforward way as follows: \( B = (\Sigma, M, Q_B, I_B, \delta_B, F_B) \), where

- \( Q_B = Q_A \times \Gamma_A \),
- \( I_B = I_A \times \{ \top \} \),
- \( F_B = F_A \times \Gamma_A \),
- \( M \) is the precedence matrix induced by the partition on \( \Sigma \):

\[
\begin{array}{c|ccc}
\Sigma_c & \Sigma_r & \Sigma_i \\
\hline
\Sigma_c & < & \equiv & < \\
\Sigma_r & > & > & > \\
\Sigma_i & > & > & > \\
# & < & < & <
\end{array}
\]

- the transition function \( \delta : Q_B \times (\Sigma \cup Q_B) \to \mathcal{P}(Q_B) \) is defined as follows, where \( q_1, q_2 \in Q_A \).

The push transition \( \delta_{\text{push}} : Q_B \to \mathcal{P}(Q_B) \) is defined

* for \( a \in \Sigma_c \), \( \delta_{\text{push}}((q_1, \gamma_1), a) = \{ (q_2, \gamma_2) \mid (q_1, a, \gamma_2, \gamma_2) \in \delta_A \} \),

---

8 Among the many equivalent definitions for VPAs we adopt here the original one in [3].
for $a \in \Sigma$, $\delta_{B_{\text{push}}}(q_1, \gamma, a) = \{q_2, \gamma\}$.
- for $a \in \Sigma$, $\delta_{B_{\text{pop}}}(q_1, \gamma) = \{q_2\}$.

The shift transition $\delta_{\text{Bshift}} : Q_B \times \Sigma \rightarrow \wp(Q_B)$ is defined
- for $a \in \Sigma$, $\delta_{\text{Bshift}}(q_1, \gamma, a) = \{q_2\}$, i.e., the
$\omega$OPBA simulates the pop move of the $\omega$BVPA by setting, as state $q_2$,
a state reached by the $\omega$BVPA while reading the return symbol $a$.

The pop transition $\delta_{\text{pop}} : Q_B \times Q_B \rightarrow \wp(Q_B)$ is defined
- for $\delta_{\text{Bpop}}(q_1, \gamma_1, q_2, \gamma_2) = \{q_1\}$, i.e., restores the state reached
by the $\omega$BVPA after its pop move.

If the original $\omega$BVPA is deterministic, so is the $\omega$OPBA obtained with the above
construction, and this yields the second relation.

**Proposition 6.2.** The following inclusion relations hold:

\[ \mathcal{L}(\omega\text{OPBEA}) \subseteq \mathcal{L}(\omega\text{OPBA}), \]
\[ \mathcal{L}(\omega\text{DOPBEA}) \subseteq \mathcal{L}(\omega\text{DOPBA}) \subseteq \mathcal{L}(\omega\text{DOPMA}) \subseteq \mathcal{L}(\omega\text{OPMA}). \]

**Proof.** The first inclusion follows trivially from the definition of $\omega$OPBA and
$\omega$OPBEA in normal form: given an $\omega$OPBEA whose set of states is partitioned into
states that are always visited with empty stack and states that are never visited with
empty stack, we can define an equivalent $\omega$OPBA that has as final states the final
states of the $\omega$OPBEA that are always visited with empty stack.

The inclusion follows similarly for the deterministic counterparts of these classes
of $\omega$OPAs, since this $\omega$OPBA is deterministic if the $\omega$OPBEA is deterministic.

About the relations involving Muller automata, $\mathcal{L}(\omega\text{DOPBA}) \subseteq \mathcal{L}(\omega\text{OPMA})$
derives from the fact that any $\omega$DOPBA $B = (\Sigma, M, Q, q_0, F, \delta)$ is equivalent to an
$\omega$DOPMA $A = (\Sigma, M, Q, q_0, T, \delta)$ whose acceptance component $T$ consists of all
subsets of $Q$ including some final state of $B$, namely, $T = \{P \subseteq Q | P \cap F \neq \emptyset\}$; the
last relation is obvious.

In the case of classical finite-state automata on infinite words, nondeterministic
Büchi automata and nondeterministic Muller automata are equivalent and define the
class of $\omega$-regular languages. Traditionally, Muller automata have been introduced to
provide an adequate acceptance mode for deterministic automata on $\omega$-words. In
fact, deterministic Büchi automata cannot recognize all $\omega$-regular languages, whereas
deterministic Muller automata are equivalent to nondeterministic Büchi ones [40].

For VPAs on infinite words, instead, the paper [4] showed that the classical de-
termination algorithm of Büchi automata into deterministic Muller automata is no
longer valid, and deterministic Muller $\omega$VPAs are strictly less powerful than nonde-
terministic Büchi $\omega$VPAs. A similar relationship holds for $\omega$OPAs too.

**Theorem 6.3.** $\mathcal{L}(\omega\text{OPBA}) = \mathcal{L}(\omega\text{OPMA}).$

**Proof.** Each $\omega$OPBA is equivalent to an $\omega$OPMA having the same underlying
OPA and acceptance component $T$ consisting of all subsets of states including some
final state of $B$ (as for their deterministic counterpart, see proof of Proposition 6.2).

Conversely, any $\omega$-language recognized by an $\omega$OPMA $A = (\Sigma, M, Q, I, T, \delta)$ can
be recognized by an $\omega$OPBA $B$ with the same precedence matrix and with $O(s^2)$
states, where $s$ is the number of states of $A$. We can assume that $T$ is a singleton.
Indeed, $L(A)$ can be expressed as

\[ L(A) = \bigcup_{T \in \mathcal{T}} L(A_T), \]  

where $A_T = (\Sigma, M, Q, I, \{T\}, \delta)$.

Since $\mathcal{L}(\omega\text{OPBA})$ is closed under union (a property that will be proved later, with
Theorem 7.6), if each language \( L(A_T) \) is accepted by an \( \omega \)OPBA, then \( L(A) \) too is accepted by an \( \omega \)OPBA.

Thus, let \( \mathcal{T} \) be the singleton \( \{T\} \). Let us build an \( \omega \)OPBA \( \mathcal{B} = \langle \Sigma, M, Q, I, F, \delta \rangle \) that accepts the same language as \( \mathcal{A} \) as follows. \( Q \) includes elements of two types: states of \( \mathcal{A} \) and states of \( (q, R) \) where \( q \in Q \) and \( R \subseteq Q \) is a set (that we informally call “box”), which will be used to test whether the run of \( \mathcal{A} \) is successful.

Intuitively, the automaton \( \mathcal{B} \) simulates \( \mathcal{A} \), reading the input string \( x \), along a sequence of states \( q \), and then guesses nondeterministically the point after which a successful run \( \rho \) of \( \mathcal{A} \) on \( x \) stops visiting the states that occur only finitely often in the run, and \( \rho \) begins to visit all and only the states in the set \( T \). After this point \( \mathcal{B} \) switches to the states of the form \( (q, R) \) and collects in \( R \) the states visited by \( \mathcal{A} \) during the run, “emptying the box” as soon as it contains exactly the set \( T \). Every time it empties the box, \( \mathcal{B} \) resumes collecting the states that \( \mathcal{A} \) will visit from that point onwards. If the final states of \( \mathcal{B} \) are defined as those ones when it collects exactly the set \( T \), then \( \mathcal{B} \) will visit infinitely often these final states iff \( \mathcal{A} \) visits all and only the states in \( T \) infinitely often.

More formally, \( \mathcal{B} \) is defined by

- \( \tilde{Q} = Q \cup (Q \times \wp(Q)) \),
- \( F = \{ (q, T) \mid q \in T \} \),
- \( \delta : \tilde{Q} \times (\Sigma \cup Q) \rightarrow \wp(Q) \), where the push function is defined by
  \[
  \delta_{\text{push}}(q, a) = \delta_{\text{push}}(q, a) \cup \{(p, \{\}) \mid p \in \delta_{\text{push}}(q, a)\} \quad \forall q \in Q, a \in \Sigma,
  \]
  \[
  \delta_{\text{push}}((q, R), a) = \begin{cases} 
  \{(p, R \cup \{\}) \mid p \in \delta_{\text{push}}(q, a)\}, & \text{if } R \neq T, \\
  \{(p, \{\}) \mid p \in \delta_{\text{push}}(q, a)\}, & \text{if } R = T
  \end{cases}
  \]
  \[\forall q, R, a \in \Sigma.\]

The shift function is defined analogously.

The pop function \( \delta_{\text{pop}} : Q \times Q \rightarrow \wp(Q) \) is defined by

- \( \delta_{\text{pop}}(q_1, q_2) = \delta_{\text{pop}}(q_1, q_2) \cup \{(p, \{\}) \mid p \in \delta_{\text{pop}}(q_1, q_2)\} \quad \forall q_1, q_2 \in Q,\]
- \( \delta_{\text{pop}}((q_1, R_1), (q_2, R_2)) = \begin{cases} 
  \{(p, R_1 \cup \{\}) \mid p \in \delta_{\text{pop}}(q_1, q_2)\}, & \text{if } R_1 \neq T, \\
  \{(p, \{\}) \mid p \in \delta_{\text{pop}}(q_1, q_2)\}, & \text{if } R_1 = T
  \end{cases}
  \]
  \[\forall q_1, q_2, R_1, R_2 \in Q.\]

First, we show that \( L(\mathcal{A}) \subseteq L(\mathcal{B}) \). Let \( x \in L(\mathcal{A}) \), and let \( \rho \) be a successful run on \( x \). There exists a finite prefix \( v \in \Sigma^* \) of \( x = vu_1u_2 \ldots \) such that the infinite path followed by \( \mathcal{A} \) after reading \( v \) (i.e., on the infinite word \( u_1u_2 \ldots \) ) visits all and only states in \( T \) infinitely often. Thus, the run \( \rho \) can be written as

\[
\rho = \langle \alpha_0 = \bot, q_0, x = vu_1u_2 \ldots \rangle \vdash^* \langle \alpha_{|v|}, q_{|v|}, u_1u_2 \ldots \rangle \vdash^+ \langle \alpha_i, q_i, u_i \ldots \rangle \vdash^+ \ldots,
\]

where \( \{q_i \mid i > |v|\} = T \) and \( q_0 \in I \). Then, there is a successful run \( \tilde{\rho} \) of \( \mathcal{B} \) on the same word, which follows singleton states of \( \mathcal{A} \) while it reads \( v \),

\[
\tilde{\rho} = \langle \beta_0 = \alpha_0 = \bot, q_0, x = vu_1u_2 \ldots \rangle \vdash^* \langle \beta_{|v|} = \alpha_{|v|}, q_{|v|}, u_1u_2 \ldots \rangle \vdash \langle \beta_{|v|+1} = \alpha_{|v|+1}, p, \{\} \rangle, \tilde{u}_1u_2 \ldots \rangle,
\]

\[
\text{and then switches to states augmented with a box: } \langle \beta_{|v|} = \alpha_{|v|}, q_{|v|}, u_1u_2 \ldots \rangle \vdash \langle \beta_{|v|+1} = \alpha_{|v|+1}, p, \{\} \rangle, \tilde{u}_1u_2 \ldots \rangle \vdash \langle \alpha_{|v|+1}, p, \tilde{u}_1u_2 \ldots \rangle \text{ and } u_1 = a\tilde{u}_1.
\]

Since after this point \( \mathcal{A} \) visits each state in \( T \) and only these states infinitely often, \( \mathcal{B} \) will reach infinitely often final states \( (q, T) \in F \), emptying infinitely often
its box as soon as it gets full, and resuming the collection of states therein with the subsequent state in the run. Conversely, $L(B) \subseteq L(A)$. Let $x \in \Sigma^\omega$ be an infinite word in $L(B)$. Define the projection $\pi : Q \cup (Q \times \mathcal{P}(Q)) \to Q$ as $\pi(q) = q$ and $\pi((q, R)) = q \forall q \in Q, R \subseteq Q$. Given a run $\tilde{\rho}$ of the automaton $B$, let $\pi(\tilde{\rho})$ be the natural extension of $\pi$ on a run. By construction, if $\tilde{\rho}$ is a run of $B$ on an $\omega$-word, then $\pi(\tilde{\rho}) = \rho$ is a run for $A$ on the same word. Now, let $\tilde{\rho}$ be a successful run for $B$ on $x$; $\rho = \pi(\tilde{\rho})$ is a run for $A$ on $x$. Since only the states augmented with a box are final states, then after a sequence (possibly empty) of singleton states initially traversed by $B$, the automaton will definitively visit only states of the form $(q, R)$ (in fact, no singleton state is reachable again from these states).

By induction on the number of final states reached by $B$ along its run, it can be proved that, for each pair of final states consecutively reached by $B$, say $(q_i, R_i)$ and $(q_{i+1}, R_{i+1})$, the portion of the run visited between them, say $\tilde{\rho}_i$, is such that the set of states reached along $\pi(\tilde{\rho}_i)$ equals exactly $T$. Finally, since final states in $\tilde{\rho}$ are visited infinitely often, the run $\pi(\tilde{\rho})$ is successful for $A$. □

6.2. Strict inclusion and incomparability results. To prove the strict inclusion and incomparability relations summarized in Figure 14, we introduce examples of $\omega$-languages, whose membership properties are summarized in Table 4.

1. For $\Sigma = \{a, b\}$, $L_{a\infty} = \{x \in \Sigma^\omega : x$ contains an infinite number of occurrences of letter $a\}$ is recognized by the $\omega$DOPBEA depicted in Figure 15.

2. $L_{a}\text{\text{- finite}} = \{x \in \Sigma^\omega : x$ contains a finite number of occurrences of $a\}$, i.e., the complement of $L_{a\infty}$, is clearly recognized by an $\omega$DOPMA and by an

![Fig. 15. $\omega$DOPBEA, with its OPM, for $L_{a\infty}$.](image-url)
\[ \{ c, r \} \] is studied in [4].

3. For \( \Sigma = \{ c, r \} \), let \( L_{\omega \text{Dyck-pr}(c, r)} \) be the language of \( \omega \)-words composed by an infinite sequence of finite-length words belonging to the Dyck language with pair \( c, r \) with possibly pending returns, i.e., letters \( r \) not matched by any previous corresponding letter \( c \). \( L_{\omega \text{Dyck-pr}(c, r)} \) is recognized by the \( \omega \)-DOPMA and the \( \omega \)-DOPBEA whose state graph is depicted in Figure 16 and with acceptance component defined, respectively, by the table \( T = \{ \{ q_0 \}, \{ q_0, q_1 \} \} \) and the set of final states \( F = \{ q_0 \} \).

4. For \( \Sigma = \{ c, r \} \), let \( L_{\text{repbsd}} \) be the language (studied in [4]) consisting of \( \omega \)-words \( x \) on \( \Sigma \) such that \( x \) has only finitely many pending calls, i.e., occurrences of letter \( c \) not matched by any subsequent corresponding letter \( r \) (\( \text{repbsd} \) stands for repeatedly bounded stack depth). \( L_{\text{repbsd}} \) is accepted by an \( \omega \)-OPBA, but cannot be accepted by any \( \omega \)-OPBEA.

Intuitively, an \( \omega \)-OPBEA accepts a word iff it reaches infinitely often a final configuration with empty stack reading the input string; however, the automaton is never able to remove all the input symbols piled on the stack since it cannot pop the pending calls interspersed among the correctly nested letters \( c \), otherwise it would either introduce conflicts in the OPM or it would not be able to verify that they are finite in number.

More formally, assume by contradiction that there is an \( \omega \)-OPBEA \( A = (\Sigma, M, Q, I, F, \delta) \) recognizing \( L_{\text{repbsd}} \). \( M \) must satisfy the following constraints since

- \( \epsilon^\omega \in L_{\text{repbsd}} \), then \( M_{\#c} = \{ \lhd \} \) and \( M_{rr} = \{ \rhd \} \);
- \( cr^\omega \in L_{\text{repbsd}}, \) then \( M_{\#c} = \{ \lhd \} \), and either \( M_{cr} = \{ \rhd \} \) or \( M_{cr} = \{ \rhd, \lhd \} \);
- \( r(cr)^\omega \in L_{\text{repbsd}} \), thus if \( c \rhd r \), \( M_{rc} = \{ \rhd \} \);
- \( c(cr)^\omega \in L_{\text{repbsd}} \), thus if \( c \rhd r \), \( M_{cc} \neq \{ \lhd \} \).

Hence, \( M \) must comply with one of the matrices \( M_1 \) or \( M_2 \) shown in Table 5.

**Table 5**

<table>
<thead>
<tr>
<th>( M_1 )</th>
<th>( c )</th>
<th>( r )</th>
<th>( M_2 )</th>
<th>( c )</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon )</td>
<td>( \lhd )</td>
<td>( \rhd )</td>
<td>( \epsilon )</td>
<td>( \lhd )</td>
<td>( \rhd )</td>
</tr>
<tr>
<td>( r )</td>
<td>( \rhd )</td>
<td>( \rhd )</td>
<td>( r )</td>
<td>( \lhd )</td>
<td>( \rhd )</td>
</tr>
<tr>
<td>#</td>
<td>( \lhd )</td>
<td>( \lhd )</td>
<td>#</td>
<td>( \lhd )</td>
<td>( \lhd )</td>
</tr>
</tbody>
</table>

\( \omega \)-OPBEA, but cannot be recognized by any \( \omega \)-DOPBA. The proof of this latter fact resembles the classical proof (see [40]) that deterministic Büchi automata are strictly weaker than nondeterministic Büchi ones.
5. For $\Sigma = \{a, b\}$, let $L_{\text{abseq}} = \{a^k b^k \mid k \geq 1\}$ and $L_{a^2 \text{abseq}} = \{x \in \Sigma^* \mid x = a^2 L_{\text{abseq}}^*\}$. Language $L_{a^2 \text{abseq}}$ is recognized by an $\omega$DOPBA, but it is not recognized by any $\omega$OPBEA (nor a fortiori by any $\omega$DOPBEA).

Indeed, words in $L_{\text{abseq}}$ can be recognized only with the OPM $M$ depicted in Figure 17: any other OPM will prevent verifying that the number of $a$’s equals that of $b$’s in subwords belonging to $L_{\text{abseq}}$. Since $a \prec a$, an $\omega$OPBEA piles up on the stack the first sequence $a^2$ of a word and cannot remove it afterwards; hence it cannot empty the stack infinitely often to accept a string in $L_{a^2 \text{abseq}}$. There is, however, an $\omega$DOPBA (and thus an $\omega$DOPMA) that recognizes such a language: it is shown in Figure 17. Notice also that $L_{a^2 \text{abseq}}$ can be recognized by an $\omega$OPBEA, with OPM $M$ and with state graph depicted in Figure 17 but with state $q_2$ instead of $q_0$ as initial state.

- If, along a run, an $\omega$OPA (or also an OPA) reaches a state with an empty stack, the subsequent suffix of the run does not depend on the transitions performed until that state.
- Since $Q$ is finite, there exist $p, q \in Q$, and an infinite set of indexes $E \subseteq \mathbb{N} \setminus \{0, 1, 2\}$ such that, for each $i \in E$, $\rho$ has a prefix: $q_0 \xrightarrow{v_i} p \xrightarrow{w_i} q$, where $v_i = c_1 e_1 \ldots i e_{i-2} c_1 e_{i-2} c_1 \ldots$ and $w_i = r c_i r$ and, given the precedence relations in $M_1$ and $M_2$, both $p$ and $q$ are reached with an empty stack, just before performing a push move while reading the letter $r$ in $w$ that follows, respectively, $v_i$ and $w_i$. For each $i \in E$, let $\rho_i$ be the finite factor of $\rho$ given by $p \xrightarrow{w_i} q$.

Let $J \subseteq E$ be the set of indexes in $E$ such that $\forall i \in J$, $\rho_i$ visits a final state with empty stack. We can build a run $\rho'$, which differs from $\rho$ in that

- for every $i \in E \setminus J$, the factor $\rho_i$ is replaced by a $\rho_j$ for some $j \in E$, with $j > i$;
- for every $i \in J$, the factor $\rho_i$ is replaced by a $\rho_j$ with $i < j \in J$ if $|J| = \infty$, or $i < j \in E$ if $|J| < \infty$.

$\rho'$ is an accepting run in $\mathcal{A}$, along which the automaton reads a word with infinitely many pending calls, which does not belong to $L_{\text{repsbd}}$, and this is a contradiction.

Furthermore, $L_{\text{repsbd}}$ is not recognizable by any $\omega$DOPMA. The proof of this fact resembles the analogous proof in [4]; indeed, that proof is essentially based on topological properties of the state graph of the automata and it is general enough to adapt to both $\omega$VPAs and $\omega$OPAs.

![Figure 17. An $\omega$DOPBA recognizing language $L_{a^2 \text{abseq}}$.](image-url)
7. Closure properties. Table 6 displays the closure properties of the various families of $\omega$-languages. In order to prove them, we first introduce some preliminary constructions in section 7.1. Then in section 7.2 we present the proofs for $L(\omega\text{OPBA})$; in particular, closure under complement and concatenation are the cases that require novel investigation techniques w.r.t. previous literature [18, 19] by referring to OPGs rather than OPAs, we present them in this part, which is where they are exploited in this paper.

### 7.1. Preliminary properties and constructions

The following constructions will be exploited to prove several closure properties. Indeed, they would be useful even to prove the same properties in the case of finite-length languages; however, since such properties have already been proved in previous literature [18, 19] by referring to OPGs rather than OPAs, we present them in this part, which is where they are exploited in this paper.

We begin by introducing the deterministic product of transition functions, defined by extending the usual construction for finite-state automata. Such a definition is meaningful when applied to automata that share the same precedence matrix, because they perform the same type of move (push/shift/pop) while reading the input word.

**Definition 7.1.** Let $Q_1$ and $Q_2$ be two disjoint sets of states of two deterministic automata sharing the same OP alphabet and let $\delta_1$ and $\delta_2$ be their transition functions. Their product state $Q$ is defined as $Q = Q_1 \times Q_2$ and their product transition function $\delta : Q \times (\Sigma \cup Q) \rightarrow Q$ is defined as follows, where $q_1, q_2, p_1, p_2 \in Q, a \in \Sigma$:

\[
\delta_{\text{push}}((q_1, q_2), a) = (\delta_{1\text{push}}(q_1, a), \delta_{2\text{push}}(q_2, a)), \\
\delta_{\text{shift}}((q_1, q_2), a) = (\delta_{1\text{shift}}(q_1, a), \delta_{2\text{shift}}(q_2, a)), \\
\delta_{\text{pop}}((q_1, q_2), (p_1, p_2)) = (\delta_{1\text{pop}}(q_1, p_1), \delta_{2\text{pop}}(q_2, p_2)).
\]

Clearly $|Q| = |Q_1| \cdot |Q_2|$.

Although this paper is not concerned with translations, we are going to need the following definition of OP Büchi $\omega$-transducers during some technical steps; other types of $\omega$-transducers could be defined similarly but are not necessary in this paper.

**Definition 7.2 (OP (Büchi) $\omega$-transducer).** An OP $\omega$-transducer is defined in the usual way as a tuple $T = (\Sigma, M, Q, I, F, O, \delta, \eta)$, where $\Sigma, M, Q, I, F$ are defined as in Definition 2.1, $O$ is a finite set of output symbols, the transition function $\delta$ and the output function $\eta$ are defined by $\langle \delta, \eta \rangle : Q \times (\Sigma \cup Q) \rightarrow \mathcal{P}_F(Q \times O^*)$, where $\mathcal{P}_F(Q \times O^*)$ denotes the set of finite subsets of $Q \times O^*$, and $\langle \delta, \eta \rangle$ can be seen as the union of three functions, $\langle \delta_{\text{shift}}, \eta_{\text{shift}} \rangle : Q \times \Sigma \rightarrow \mathcal{P}_F(Q \times O^*)$, $\langle \delta_{\text{push}}, \eta_{\text{push}} \rangle : Q \times \Sigma \rightarrow \mathcal{P}_F(Q \times O^*)$, and $\langle \delta_{\text{pop}}, \eta_{\text{pop}} \rangle : Q \times Q \rightarrow \mathcal{P}_F(Q \times O^*)$.

### Table 6

<table>
<thead>
<tr>
<th>Closure properties of families of $\omega$-languages. ($L_1 \cdot L_2$ denotes the concatenation of a language of finite-length words $L_1$ with an $\omega$-language $L_2$).</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{L}(\omega\text{DOPBA})$</td>
</tr>
<tr>
<td>Intersection</td>
</tr>
<tr>
<td>Union</td>
</tr>
<tr>
<td>Complement</td>
</tr>
<tr>
<td>$L_1 \cdot L_2$</td>
</tr>
</tbody>
</table>
A configuration of the $\omega$-transducer is denoted $\langle \beta, q, w \rangle \downarrow z$, where $C = \langle \beta, q, w \rangle$ is the configuration of the underlying $\omega$OPBA and the string after $\downarrow$ represents the output of the automaton in the configuration. The transition relation $\vdash$ is naturally extended from $\omega$OPBAs, by concatenating the output symbols produced at each move with those generated in the previous moves. Runs and acceptance by the transducer are defined as in the corresponding $\omega$OPBA.

The transduction $\tau(x), x \in \Sigma^\omega$, generated by $T$ is the set of $\omega$-strings produced during its nondeterministic successful runs over $x$.

The next statement is propaedeutic to many constructive proofs of closure properties, where the operands are in general OPAs with compatible but not identical matrices, and the result’s matrix must often be the union of the two original ones. If $A$ is an OPA with precedence matrix $M$ and $M' \supseteq M$, then clearly $A$ works also over $M'$ but the language recognized by $A$ over $M'$ is not necessarily the same, since the presence of precedence relations in $M'$ that are not included in $M$ may allow for successful runs on some words that are, instead, not successful in the original OPA. The next statement proves, however, that the precedence matrix of an OPA can always be extended (up to completion), provided that conflict freedom is preserved, without affecting the recognized language.

**Statement 7.1 (extended matrix normal form).** Let $A = \langle \Sigma, M, Q, I, F, \delta \rangle$ be an OPA (over finite-length or omega words) with $|Q| = s$. For any conflict-free OPM $M' \supseteq M$, there exists an OPA with OPM $M'$ that recognizes the same language as $A$ and has $O(|\Sigma|^2 s)$ states.

**Proof.** First consider finite-length words. The new OPA $A' = \langle \Sigma, M', Q', I', F', \delta' \rangle$ is derived from $A$ in the following way:

- $Q' = \Sigma \times Q \times \Sigma$, where $\Sigma = (\Sigma \cup \{\#\})$, i.e., the first component of a state is the lookback symbol, the second component is a state of $A$, and the third component is the lookahead symbol;
- $I' = \{\#\} \times I \times \{a \in \Sigma \mid M_{#a} \neq \emptyset\}$;
- $F' = \{\#\} \times F \times \{\#\}$;
- $\delta' : Q' \times (\Sigma \cup Q') \to \mathcal{P}(Q')$ is the transition function defined as follows.

Let $a \in \Sigma, b \in Q, q \in Q$. The push transition $\delta'_\text{push} : Q' \times \Sigma \to \mathcal{P}(Q')$ is defined by

$$
\delta'_\text{push}(\langle a, q, b \rangle, b) = \{\langle b, p, c \rangle \mid p \in \delta_\text{push}(q, b) \land M_{ab} = \{\ immature \} \land M_{bc} \neq \emptyset\}.
$$

The shift transition $\delta'_\text{shift} : Q' \times \Sigma \to \mathcal{P}(Q')$ is defined analogously:

$$
\delta'_\text{shift}(\langle a, q, b \rangle, b) = \{\langle b, p, c \rangle \mid p \in \delta_\text{shift}(q, b) \land M_{ab} = \{\ immature \} \land M_{bc} \neq \emptyset\}.
$$

The pop transition $\delta'_\text{pop} : Q' \times Q' \to \mathcal{P}(Q')$ is defined by

$$
\delta'_\text{pop}(\langle a_1, q_1, a_2 \rangle, \langle b_1, q_2, b_2 \rangle) = \left\{ \langle b_1, q_3, a_2 \rangle : q_3 \in \delta_\text{pop}(q_1, q_2) \land M_{a_1 a_2} = \{\ immature \} \land M_{b_1 a_2} \neq \emptyset \right\},
$$

where $a_1, b_2 \in \Sigma, a_2, b_1 \in \Sigma, q_1, q_2 \in Q$.

Clearly, the OPA $A'$ has OPM $M'$ and accepts the same language as $A$.

This construction can be naturally extended to $\omega$OPBAs: in particular, for $\omega$OPBA the set of final states of $A'$ is $F' = \Sigma \times F \times \Sigma$, i.e., a run of $A'$ is accepting if it visits infinitely often final states of $A$, independently of the lookback and the lookahead symbols considered for these states. For $\omega$OPBEA this acceptance component may be further refined as $F' = \{\#\} \times F \times \Sigma$. For $\omega$OPMA, $T' = \{t \mid t = A_1 \times S \times A_2, S \in T, A_1 \subseteq \Sigma, A_2 \subseteq \Sigma\}$, where $T \subseteq \mathcal{P}(Q)$ is the table of $A$. Furthermore, the transformation preserves determinism. [1]
7.1.1. OPA's version without # as lookahead. In this section we illustrate
a new version of OPAs that do not rely on the end marker # for the recognition of a
finite-length word.

The new model is defined by slightly modifying the semantics of the transition
relation and of the acceptance condition of original OPAs, in such a way that a string
is accepted by an automaton if it reaches a final state right at the end of the parsing
of the whole word, and does not perform any pop move determined by the ending
delimiter # to empty the stack; thus the automaton stops just after having pushed on
the stack (or updated the top of the stack symbol with) the last symbol of the string.

In this alternative characterization of OPAs, the semantics of the transition re-
lation differs from the classical definition in that, once a configuration with the end
marker as lookahead is reached, the computation cannot evolve in any subsequent con-
figuration, i.e., a pop move \( C_1 \vdash_\gamma C_2 \) with \( C_1 = (\Pi[a, p], q, x#) \) is performed only
if \( x \neq \varepsilon \) (where symbol \( \vdash_\gamma \) denotes a move according to this variation of the semantics
of the transition relation). The language accepted by the automaton according to this
new semantics (denoted as \( L_\gamma \)) is the set of words

\[
L_\gamma(A) = \{ x | \langle \bot, q_1, x# \rangle \vdash_\gamma^* \langle \bot, q_f, \#, \rangle, q_1 \in I, q_f \in F, \gamma \in \Gamma^* \}.
\]

This new version of the automaton, called \#-lookahead OPA (\#OPA) is closer to
the traditional acceptance criterion of general pushdown automata; we emphasize,
however, that, unlike normal acceptance by the final state of a pushdown automaton,
which can perform a number of \( \varepsilon \)-moves after reaching the end of a string and accepts
it if just one of the visited states is final, this type of automaton cannot perform any
(pop, i.e., \( \varepsilon \)-) move when it reaches the end of the input string. The following lemmas
(Lemmas 7.3 and 7.4) prove the equivalence between the original version of OPAs and
the new one.\(^9\)

**Lemma 7.3.** Let \( A_1 \) be a nondeterministic OPA defined on an OP alphabet \((\Sigma, M)\)
with \( s \) states. Then there exists a nondeterministic \#OPA \( A_2 \) on \((\Sigma, M)\) and \( O(s^2) \)
states such that \( L(A_1) = L_\gamma(A_2) \).

We first explain informally the rationale of the simulation of \( A_1 \) by \( A_2 \), with the
aid of an example; then we formally define its construction and prove their equivalence.

Consider a word of finite length \( w \) compatible with \( M \): the string \( \#w \) can be
factored in a unique way as a sequence of bodies of chains and pending letters as

\[
\# w = \# a_1a_2a_2 \ldots a_na_n,
\]

where \( a_{i-1}[w_i]a_i \) are maximal chains and each \( w_i \) can be possibly missing with \( a_0 = \# \)
and \( \forall i : 1 \leq i \leq n-1, a_i \leq a_{i+1} \) or \( a_i = a_{i+1} \). Let \( i_j \in \{1, 2, \ldots, n\} \), \( 1 \leq j \leq k, k \geq 1 \),
be indexes such that

\[
\#a_{i_1} = a_1 \sqcup \cdots \sqcup a_{i_2-1} \sqcup a_{i_2} \sqcup \cdots \sqcup a_{i_3-1} \sqcup a_{i_3} \sqcup \cdots \sqcup a_{i_{k-1}} \sqcup a_{i_k} \sqcup a_{i_{k+1}} \ldots \sqcup a_n.
\]

When reading \( w \), the symbols of the string are progressively put on the stack, either
by a push move or by a shift move and, whenever a chain \( w_i \) is recognized, the symbol
on the top of the stack is popped. Hence, after reading \( w \) the stack contains only the
symbols \( \# a_{i_{1-1}} a_{i_{2-1}} \ldots a_n \) that are the ending symbols of the open chains in the
sequence (7.1).

---

\(^9\)Only Lemma 7.3 will be used in Part 4.3 of this paper but we include both for completeness
and possible further exploitation.
When \( w \) is read by a standard OPA, the automaton performs a series of pop moves at the end of the string due to the presence of the end delimiter \( # \). These moves progressively empty the stack. The run is accepting if it leads to a final state after all pop moves.

A nondeterministic automaton that, unlike standard OPAs, does not resort to the end delimiter \( # \) for the recognition of a string must guess nondeterministically the ending point of each open chain and guess how, in an accepting run, the states in these points would be updated if the final pop moves were progressively performed. The automaton must behave as if, at the same time, it simulates two snapshots of the accepting run of a standard OPA: a move during the reading of the input, and a step during the final pop transitions which will later on empty the stack, leading to a final state. To this aim, the states of a standard OPA are augmented with an additional component.

A \( \mathcal{SS}#OPA \) equivalent to a given OPA \( \mathcal{A} \), thus may be defined so that, after reading each prefix of a word, it reaches a final state whenever, if the word were completed in that point with \( # \), \( \mathcal{A} \) could reach an accepting state with a sequence of pop moves. In this way, \( \mathcal{A} \) can guess in advance which words may eventually lead to an accepting state of \( \mathcal{A} \). Before going into the details of the construction, the following example illustrates the above intuitive description.

**Example 9.** We refer to the computation of the OPA in Example 3. Consider the input word of this computation without the end marker \( # \). The sequence of pending letters in the input word corresponds to three open chains, according to (7.1), with starting symbols \(+\), \(\times\), \(\llparenthesis\), respectively.

Figure 18 shows the configuration just before looking ahead at the symbol \( # \). The states depicted within a box are those placeholders that an equivalent \( \mathcal{SS}#OPA \) should fill up to guess in advance the last pop moves \( q_3 = q_3 \overset{q_3}{\rightarrow} q_3 \overset{q_3}{\rightarrow} q_3 \overset{q_3}{\rightarrow} q_3 \in F_1 \) of the accepting run. The corresponding configuration of the \( \mathcal{SS}#OPA \) is depicted in Figure 19.

**Proof of Lemma 7.3.** Let \( \mathcal{A}_1 = (\Sigma, M, Q_1, I_1, F_1, \delta_1) \) and define \( \mathcal{A}_2 = (\Sigma, M, Q_2, I_2, F_2, \delta_2) \) as follows.
- \( Q_2 = \{B, Z, U\} \times Q_1 \times Q_1 \).

Hence, a state \( (x, q, p) \) of \( \mathcal{A}_2 \) is a tuple whose first component denotes a nondeterministic guess for the next input symbol to be read, i.e., whether it is a pending letter which is the initial symbol of an open chain (Z), or a
consider an accepting run of the OPA that is equivalent to the OPA of Example 3.

Table 7

<table>
<thead>
<tr>
<th>Stack</th>
<th>State</th>
<th>Current Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>\item n \cdot (B, q_0, q_3)</td>
<td>(B, q_0, q_3)</td>
<td>\text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(B, q_0, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(B, q_1, q_2) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
<tr>
<td>+</td>
<td>(Z, q_1, q_3)</td>
<td>(Z, q_1, q_3) \text{n + n \times {n + n} #}</td>
</tr>
</tbody>
</table>

pending letter within an open chain other than the first one (U), or a symbol within a maximal chain (B). The second component of a state represents the current state \(q\) in \(A_1\). To illustrate the meaning of the last component, consider an accepting run of \(A_1\) and let \(q\) be its current state just before a push move to be performed when reading the first symbol of an open chain; also, let \(r\) be the state reached by such a push move and \(s\) be the state of the automaton when the stack element pushed by this move (possibly updated by subsequent shifts) is going to be popped leading to a state \(p\). Then, in the same position of the corresponding run of \(A_2\), the current state would be \((Z, q, p) \in Q_2\) and state \(\langle x, r, s \rangle \in Q_2\) will be reached by \(A_2\) (\(x\) being nondeterministically any one of \(B, Z, U\)); in other words, the last component \(p\) represents a guess about the state that will be reached in \(A_1\) when the stack element pushed by this move will be popped. Hence we can consider only states \((Z, q, p) \in Q_2\) such that \(s \overset{q}{\Rightarrow} p\) in \(A_1\) for some \(s \in Q_1\). In all the other positions the last component is simply propagated.

For instance, Table 7 shows an accepting run on the word \(n + n \times \{n + n\}\) of a \(\mathcal{HPA}\) that is equivalent to the OPA of Example 3. Note that before reading the \(\{\), which is the beginning of an open chain, the automaton is in the state \((Z, q, q_3)\) and then moves to \((B, q_2, q_3)\) guessing the state that is reached by the pop move that occurs in the corresponding run of the OPA after reading the \(\}\). Before reading the second \(n\), which is the body of a maximal chain, instead, the automaton is in state \((B, q_0, q_3)\) and, after popping \(n\) from the stack, moves to \((Z, q_1, q_3)\) since the following \(\times\) is the beginning of an open chain.

- \(I_2 = \{\langle x, q, q_3 \rangle | x \in \{Z, B\}, q \in I_1, q_3 \in F_1\}\).
- \(F_2 = \{\langle Z, q, q \rangle | q \in Q_1\}\).
- The transition function is defined as the union of three functions. The push transition function \(\delta_{2\text{push}} : Q_2 \times \Sigma \rightarrow 2(Q_2)\) is defined as follows, where \(p, q, r, s \in Q_1, a \in \Sigma\).
  - **Pending letter at the beginning of an open chain:**
    \[
    \delta_{2\text{push}}\left((Z, q, p), a\right) = \left\{\langle x, r, s\rangle | x \in \{B, Z, U\}, r \in \delta_{1\text{push}}(q), s \overset{a}{\Rightarrow} p\right\}.\]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
V. Lonati, D. Mandrioli, F. Panella, M. Pradella

- Symbol of a maximal chain:

\[ \delta_{2\text{push}}(\langle B, q, p \rangle, a) = \{ \langle B, r, p \rangle \mid r \in \delta_{1\text{push}}(q, a) \} . \]

The shift transition function \( \delta_{2\text{shift}} : Q_2 \times \Sigma \rightarrow \mathcal{P}(Q_2) \) is defined as follows:

- Pending letter within an open chain:

\[ \delta_{2\text{shift}}(\langle U, q, p \rangle, a) = \{ \langle x, r, p \rangle \mid x \in \{ B, Z, U \}, r \in \delta_{1\text{shift}}(q, a) \} . \]

- Symbol of a maximal chain:

\[ \delta_{2\text{shift}}(\langle B, q, p \rangle, a) = \{ \langle B, r, p \rangle \mid r \in \delta_{1\text{shift}}(q, a) \} . \]

Notice that the second component of the states computed by \( \delta_{2\text{push}} \) and \( \delta_{2\text{shift}} \) is independent of the first component of the starting state.

The pop transition function \( \delta_{2\text{pop}} : Q_2 \times Q_2 \rightarrow \mathcal{P}(Q_2) \) can be executed only within a maximal chain since there is no pop determined by the ending delimiter:

\[ \delta_{2\text{pop}}(\langle B, q, s \rangle, \langle B, p, s \rangle) = \{ \langle x, r, s \rangle \mid x \in \{ B, Z, U \}, q \xrightarrow{p} r \text{ in } A_1 \} . \]

All other moves lead to an error state.

Let us prove first \( L(A_1) \subseteq L(A_2) \). Consider a word \( w \in L(A_1) \). Then there exists a support \( q \xrightarrow{w} q' \) in \( A_1 \) with \( q \in I_1 \) and \( q' \in F_1 \). If \( w = w_1 a_1 w_2 a_2 \ldots w_n a_n \), where \( a_i \) are pending letters and \( w_i \) are maximal chains, let \( k \) be the number of open chains determined by the sequence of pending letters in \( w \) according to the structure (7.1), and let \( a_{i_1}, a_{i_2}, \ldots, a_{i_k} \) be their initial symbols. Also, for every \( i = 2, \ldots, n \), let \( t(i) \) be the greatest index \( t \) such that \( i_t < i \), i.e., \( a_i \) is within the \( t(i) \)th open chain beginning with \( a_{i_t(i)} \). In particular, for \( i = n \), if \( a_{n-1} < a_n \) then \( i_n = n \), otherwise \( i(n) = k \). As a notational convention, denote by \( \xrightarrow{} \) a move that can be either a push or a shift.

Then the above support for \( w \) can be decomposed as

\[
q = \tilde{q}_0 \xrightarrow{w_1} q_1 \xrightarrow{a_1} \tilde{q}_1 \xrightarrow{w_2} q_2 \xrightarrow{a_2} \cdots \xrightarrow{w_n} q_n \xrightarrow{a_n} \tilde{q}_n = p_k ,
\]

\[
\tilde{q}_n = p_k \quad q_{k-1} \xrightarrow{q_{k-1}} p_{k-2} \xrightarrow{\cdots} p_2 \xrightarrow{q_2} p_1 \xrightarrow{q_1} p_0 = q' ,
\]

where \( q_i = \tilde{q}_i - 1 \) if \( w_i = \varepsilon \) for \( i = 1, 2, \ldots, n \). Notice that, for every \( t \), \( q_{i_t} \) is the state reached in this path before the push move that pushes symbol \( a_{i_t} \) on the stack; moreover, when the last symbol in the open chain beginning with \( a_{i_t} \) is to be popped, the current state is \( p_t \) and then the symbol on the top of the stack (whose state component is \( q_{i_t} \)) is removed and \( A_1 \) moves to state \( p_{t-1} \).

Starting with state \( \langle Z, q_1, p_0 \rangle \) if \( w_1 = \varepsilon \) or with \( \langle B, \tilde{q}_0, p_0 \rangle \) if \( w_1 \neq \varepsilon \), an accepting computation of \( A_2 \) can be built on the basis of the following facts:

- Since \( A_1 \) performs \( q_1 \xrightarrow{a_1} \tilde{q}_1 \) and \( p_1 \xrightarrow{q_1} p_0 \), then \( \delta_{2\text{push}}(\langle Z, q_1, p_0 \rangle, a_1) \ni \langle x, q_1, p_1 \rangle \) for \( x \in \{ B, Z, U \} \). This is a push move that can be applied at the beginning of the first open chain, \( a_1 \), where \( p_1 \) is the guess about the state that will be reached before the stack symbol pushed on the stack by this move will be popped.

- In general, for every \( t \), since \( A_1 \) executes \( q_{i_t} \xrightarrow{a_{i_t}} \tilde{q}_{i_t} \) and \( p_t \xrightarrow{q_{i_t}} p_{t-1} \), then \( \delta_2(\langle Z, q_{i_t}, p_{t-1} \rangle, a_{i_t}) \ni \langle x, q_{i_t}, p_t \rangle \) for \( x \in \{ B, Z, U \} \). This is a push move that
can be applied at the beginning of the \( t \)th open chain, i.e., when reading \( a_{i_1} \),
where \( p_i \) is the guess about the state that will be reached before the stack symbol with the last letter of the chain will be popped. In particular, if \( i_k = n \), we can reach state \( \langle Z, q_n, p_k \rangle \) which is final in \( \mathcal{A}_2 \) since \( q_n = p_k \).
- For every maximal chain \( w_i \) of \( w \) (with \( i \geq 2 \)) consider its support \( \tilde{q}_{i-1} \sim q_i \) in the sequence (7.2). Then in \( \mathcal{A}_2 \) we have a sequence of moves starting from a state \( \langle B, \tilde{q}_{i-1}, p_{(i)} \rangle \) and reading \( w_i \), that ends in \( \langle x, q_i, p_{(i)} \rangle \), where \( x \in \{ U, Z \} \). Notice that the last component of the states does not change because we are within a maximal chain. During the reading of \( w_i \), the last component is equal to \( p_{(i)} \), as guessed by the push move at the beginning of the current open chain.
- For every \( i \not\in \{ i_1, i_2, \ldots, i_k \} \), since \( \delta_{\text{shift}}(q_i, a_t) \ni \tilde{q}_i \), then \( \delta_{\text{shift}}(\langle U, q_i, p_{(i)} \rangle, a_t) \) contains \( \langle x, \tilde{q}_i, p_{(i)} \rangle \), for \( x \in \{ B, Z, U \} \). In particular, if \( n \neq i_k \), then \( t(n) = k \) and for \( i = n \) we can reach state \( \langle Z, q_n, p_k \rangle \) which is final in \( \mathcal{A}_2 \), since \( q_n = p_k \).

Thus, by composing in the right order the previous moves, one can obtain an accepting computation for \( w \) in \( \mathcal{A}_2 \).

Conversely, to prove that \( \mathcal{L}_{\text{shift}}(\mathcal{A}_2) \subseteq \mathcal{L}(\mathcal{A}_1) \), consider a word \( w \in \mathcal{L}_{\text{shift}}(\mathcal{A}_2) \). This means that there exists a successful run of \( \mathcal{A}_2 \) on \( w \). Let \( w \) be factorized as above; then the accepting run for \( w \) can be decomposed as

\[
\pi_0 \sim w_1 \pi_1 \xrightarrow{a_1} \pi_2 \sim w_2 \pi_2 \xrightarrow{a_2} \cdots \pi_i \sim w_i \pi_i \xrightarrow{a_i} \cdots \pi_n \sim w_n \pi_n,\
\]

where \( \pi_i, \pi_0 \in Q_2 \), \( \pi_i = \pi_{i-1} \) if \( w_i = \varepsilon \), \( \pi_0 \in I_2 \), and \( \pi_n \in F_2 \). By projecting this path on the second component of states \( \pi_i \) and \( \pi_0 \) (let them, resp., be \( p_i \) and \( r_i \) in \( Q_1 \)), we obtain a path in \( \mathcal{A}_1 \) labeled by \( w \). This path is not accepting because there are symbols left on the stack that need to be popped, but we can complete this path arguing by induction on the structure of maximal chains according to the definition of \( \delta_2 \). Precisely, one can verify that \( Q_1 \) contains suitable states \( p_i \) (for \( 0 \leq i \leq n \)), \( r_i \) (for \( 1 \leq i \leq n \)), \( s_t \) (for \( 1 \leq t \leq k \)), with \( r_i = p_{i-1} \) whenever \( w_i = \varepsilon \), such that the following facts hold:

- \( \pi_0 \in I_2 \), hence \( \pi_0 = \langle x_0, p_0, s_0 \rangle \) with \( p_0 \in I_1 \) and \( s_0 \in F_1 \); \( x_0 = B \) if \( w_1 \neq \varepsilon \), otherwise \( x_0 = Z \).
- \( \pi_0 \sim p_1 \) in \( \mathcal{A}_2 \) implies that the last component of state \( \pi_0 \) is propagated through chain \( w_1 \) without change; hence \( \pi_1 = \langle Z, r_1, s_0 \rangle \) with \( p_0 \sim r_1 \) in \( \mathcal{A}_1 \).
- \( p_1 \xrightarrow{a_1} \pi_1 \) is a push move of \( \mathcal{A}_2 \) at the beginning of an open chain, and this implies that the last component of \( \pi_1 \) is a guess on the state from which \( \mathcal{A}_1 \) would perform the corresponding pop, so that \( \pi_1 = \langle x_1, p_1, s_1 \rangle \) with \( r_1 \xrightarrow{a_1} p_1 \) and \( s_1 \xrightarrow{\varepsilon} s_0 \) in \( \mathcal{A}_1 \); the first component is \( x_1 = B \) if \( w_2 \neq \varepsilon \) otherwise \( x_1 \) equals \( Z \) or \( U \) according to whether \( a_2 \) starts an open chain or not, respectively.
- The pop moves within \( \pi_i \xrightarrow{a_{i+1}} \pi_{i+1} \) for \( 1 \leq i < i_2 \), and the shift moves within an open chain \( p_i \xrightarrow{a_{i}} \pi_i \) for \( 1 < i < i_2 \) propagate the last component without changing it. Hence \( \pi_i = \langle U, r_i, s_1 \rangle \) and \( \pi_i = \langle x_i, p_i, s_1 \rangle \) with \( p_{i-1} \xrightarrow{a_{i}} r_i \xrightarrow{a_{i}} p_i \) in \( \mathcal{A}_1 \). The first component is \( x_i = B \) if \( w_i \neq \varepsilon \), otherwise \( x_i = Z \) for \( i = i_2 - 1 \), and \( x_i = U \) in the other cases.
- \( p_{i_2} \xrightarrow{a_{i_2}} \pi_{i_2} \) is a push move of \( \mathcal{A}_2 \) at the beginning of an open chain, and this implies that the last component of \( \pi_{i_2} \) is a guess on the state from which \( \mathcal{A}_1 \) would perform the corresponding pop, so that \( \pi_2 = \langle x_{i_2}, p_{i_2}, s_2 \rangle \) with \( r_{i_2} \xrightarrow{a_{i_2}} p_{i_2} \) and \( s_2 \xrightarrow{\varepsilon} s_1 \) in \( \mathcal{A}_1 \). The first component is \( x_{i_2} = B \) if \( w_{i_2} \neq \varepsilon \)
otherwise \( x_1 \) equals \( Z \) or \( U \) according to whether \( a_{i_2} + 1 \) begins an open chain or not, respectively.

- Similarly for the following moves in the run.

In general, we get

\[
\rho_i = \langle y_i, r_i, s_{t(i)} \rangle \quad \text{for every } i = 1, 2, \ldots, n,
\]

\[
\pi_i = \langle x_i, p_i, s_{t(i)} \rangle \quad \text{for every } i \notin \{i_1, i_2, \ldots, i_k\},
\]

\[
\pi_{i_k} = \langle x_k, p_{i_k}, s_t \rangle \quad \text{for every } t = 1, 2, \ldots, k
\]

with \( r_i \xrightarrow{a_i} p_i, s_t \xrightarrow{r_{i_k}} s_{t-1}, p_{i-1} \xrightarrow{u_i} r_i \in \mathcal{A}_1 \),

\[
\text{and } y_i \in \{Z, U\}, x_i \in \{B, Z, U\} \quad \text{for every } i \text{ and } t.
\]

For \( i = n \) we have \( n = i_k \) or \( t(n) = k \), hence \( \pi_n = \langle x_n, p_n, s_k \rangle \), and \( p_n = s_k \) and \( x_n = Z \) since \( \pi_n \in F_2 \). Thus, in \( \mathcal{A}_1 \) there is an accepting run

\[
\begin{align*}
I_1 \ni p_0 \xrightarrow{u_1} r_1 \xrightarrow{a_1} p_1 \xrightarrow{u_2} r_2 \xrightarrow{a_2} \cdots \xrightarrow{u_{i_k}} r_{i_k} \xrightarrow{a_{i_k}} p_{i_k} \xrightarrow{u_{i+1}} \cdots \xrightarrow{u_k} r_k \xrightarrow{a_k} p_k = s_k, \\
p_n = s_k \xrightarrow{r_k} s_{k-1} \xrightarrow{r_{k-1}} s_{k-2} \xrightarrow{r_{k-2}} \cdots \xrightarrow{r_2} s_2 \xrightarrow{r_1} s_1 \xrightarrow{r_1} s_0 \in F_1. 
\end{align*}
\]

The next lemma completes the proof of equivalence between OPAs and \( \mathbb{X}\text{OPAs}.

\textbf{Lemma 7.4.} Let \( \mathcal{A}_2 \) be a nondeterministic \( \mathbb{X}\text{OPA} \) defined on an OP alphabet \( (\Sigma, M) \) with \( s \) states. Then there exists a nondeterministic OPA \( \mathcal{A}_1 \) on \( (\Sigma, M) \) and \( O(|\Sigma|s) \) states, such that \( L(\mathcal{A}_1) = L_{\mathbb{X}}(\mathcal{A}_2) \).

\textbf{Proof.} Let \( \mathcal{A}_2 \) be \( (\Sigma, M, Q, I, F, \delta) \) and consider, first, an equivalent form of \( \mathcal{A}_2 \), where all states are enriched with a lookahead symbol and no final state is reached by a pop edge: \( \mathcal{A}_2 = (\Sigma, M, Q_2, I_2, F_2, \delta_2) \), where

- \( Q_2 = Q \times \Sigma \), where \( \Sigma = (\Sigma \cup \{\#\}) \), i.e., the first component of a state is a state of \( \mathcal{A}_2 \) and the second component of the state is the lookahead symbol;
- \( I_2 = I \times \{a \in \Sigma \mid M_{\#a} \neq \emptyset\} \) is the set of initial states of \( \mathcal{A}_2 \);
- \( F_2 = F \times \{\#\} \);
- the transition function \( \delta_2 : Q_2 \times (\Sigma \cup Q_2) \to \mathcal{P}(Q_2) \) is defined in the following natural way, where \( a, b \in \Sigma, p, q, r \in Q \):
  - \( \delta_2\text{push}((p, a), a) = \{\langle q, b \rangle \mid q \in \delta_2\text{push}(p, a) \land M_{ab} \neq \emptyset\} \),
  - \( \delta_2\text{shift}((p, a), a) = \{\langle q, b \rangle \mid q \in \delta_2\text{shift}(p, a) \land M_{ab} \neq \emptyset\} \),
  - \( \delta_2\text{pop}((p, a), \langle q, b \rangle) = \{\langle r, a \rangle \mid r \in \delta_2\text{pop}(p, q)\} \setminus F_2 \).

It is easy to see that \( L_{\mathbb{X}}(\mathcal{A}_2) = L_{\mathbb{X}}(\mathcal{A}_2) \). Furthermore, the final states of \( \tilde{\mathcal{A}}_2 \) cannot be reached by pop edges: in fact, these pop transitions cannot be performed by a \( \mathbb{X}\text{OPA} \) according to the semantics of the transition relation \( \rightharpoonup_{\mathbb{X}} \), since it stops a computation right before reading the delimiter \( \# \), when the parsing of the word ends.

Thus, we build, without loss of generality, an OPA \( \mathcal{A}_1 \) equivalent to the \( \mathbb{X}\text{OPA} \) \( \tilde{\mathcal{A}}_2 \). \( \mathcal{A}_1 = (\Sigma, M, Q_1, I_1, F_1, \delta_1) \) has only one final state, reachable through a pop edge by all final states of \( \tilde{\mathcal{A}}_2 \). Its role is to let \( \mathcal{A}_1 \) empty the stack after reading a word that is accepted by \( \tilde{\mathcal{A}}_2 \).

- \( Q_1 = Q_2 \cup \{q_{\text{accept}}\} \).
- \( I_1 = I_2 \cup \{q_{\text{accept}}\} \) if \( I_2 \cap F_2 \neq \emptyset \); otherwise \( I_1 = I_2 \).
- \( F_1 = \{q_{\text{accept}}\} \).
- The transition function \( \delta_1 \) equals \( \delta_2 \) on all states in \( Q_2 \); in addition, \( \mathcal{A}_1 \) has departing pop edges from the final states in \( F_2 \) to \( q_{\text{accept}} \) and \( q_{\text{accept}} \) has no outgoing push/shift edge but only self-loops pop edges.
- The push transition function \( \delta_{1\text{push}} : Q_1 \times \Sigma \to \mathcal{P}(Q_1) \) is defined as \( \delta_{1\text{push}}(q, c) = \delta_{2\text{push}}(q, c) \forall q \in Q_2, c \in \Sigma \). The shift function is defined analogously.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
The pop transition \( \delta_{\text{pop}} : Q_1 \times Q_1 \rightarrow \mathcal{P}(Q_1) \) is defined by
\[
\delta_{\text{pop}}(q, p) = \delta_{2\text{pop}}(q, p), \forall q, p \in Q_2,
\]
\[
\delta_{\text{pop}}(q, p) = q_{\text{accept}}, \forall q \in (F_2 \cup \{q_{\text{accept}}\}), p \in Q_2.
\]

We now show that \( L(A_1) = L_{\text{opt}}(\tilde{A}_2) \).

\( L(A_1) \subseteq L_{\text{opt}}(\tilde{A}_2) \): in fact, if the OPA \( A_1 \) recognizes a word, then it is either the empty word and thus \( q_{\text{accept}} \in I_1 \) and also \( \tilde{A}_2 \) has a successful run on it, or \( A_1 \) recognizes a word \( w \neq \varepsilon \) and there exists a run \( \sigma \) of \( A_1 \) which ends in the final state \( q_{\text{accept}} \) with empty stack. Notice that \( q_{\text{accept}} \) is reached by a pop move from a state in \( F_2 \), say \( q_f \in F_2 \):
\[
\sigma : q_0 \in I_2 \xrightarrow{w} q_f \Rightarrow q_{\text{accept}}(p \in Q_1)^{*},
\]
and \( q_f \) itself is reached exactly when the reading of \( w \) is finished, since, as said before, a state in \( F_2 \) cannot be reached by pop moves. This condition is necessary to avoid the presence of sequences of pop moves from nonaccepting states toward final states. Then the path from \( q_0 \) to \( q_f \), which traverses the same states and edges as \( \sigma \), represents a run of \( \tilde{A}_2 \) which ends in a final state \( q_f \) right after the reading of the whole word, thus accepting \( w \). Conversely, the relation \( L(A_1) \supseteq L_{\text{opt}}(\tilde{A}_2) \) derives easily from the fact that, if \( \tilde{A}_2 \) accepts a word along a successful run, then \( A_1 \) recognizes the word along the same run, possibly emptying the stack in the final state \( q_{\text{accept}} \).

Remark 2. With some further effort—and a further exponential leap in the automaton’s size—a deterministic version of this \( \text{OPA} \) could also be built. We did not include it here, however, since the \( \text{OPA} \) construction will be applied only in this part to prove the closure w.r.t. concatenation with finite-length languages of \( \omega \text{OPLs} \): we will see that such a closure holds only for nondeterministic automata.

7.2. Closure properties and emptiness problem for class \( \mathcal{L}(\omega \text{OPBA}) \).

\( \mathcal{L}(\omega \text{OPBA}) \) enjoys all closure and decidability properties suitable for model checking. Precisely, the emptiness problem is decidable for OPAs in polynomial time because they can be interpreted as pushdown automata on infinite-length words, e.g., [13] shows an algorithm that decides the alternation-free modal \( \mu \)-calculus for context-free processes, with linear complexity in the size of the system’s representation.

The following theorems state that \( \mathcal{L}(\omega \text{OPBA}) \) is a Boolean algebra closed under concatenation.

**Theorem 7.5** (\( \mathcal{L}(\omega \text{OPBA}) \) is closed under intersection). Let \( L_1 \) and \( L_2 \) be \( \omega \)-languages recognized by two \( \omega \text{OPBA} \) defined over the same alphabet \( \Sigma \), with compatible precedence matrices \( M_1 \) and \( M_2 \) and with \( s_1 \) and \( s_2 \) states, respectively. Then \( L = L_1 \cap L_2 \) is recognizable by an \( \omega \text{OPBA} \) with OPM \( M = M_1 \cap M_2 \) and \( O(s_1, s_2) \) states.

**Proof.** Let \( A_1 = (\Sigma, M_1, Q_1, I_1, F_1, \delta_1) \) and \( A_2 = (\Sigma, M_2, Q_2, I_2, F_2, \delta_2) \) be two \( \omega \text{OPBA} \) with \( L(A_1) = L_1 \) and \( L(A_2) = L_2 \) and with compatible precedence matrices \( M_1 \) and \( M_2 \). Suppose, without loss of generality, that \( Q_1 \) and \( Q_2 \) are disjoint and do not contain \( \{0, 1, 2\} \).

First, observe that, the two OPMs being compatible, at each move either the two automata perform the same type of move (push/shift/pop), or at least one of them stops without accepting since its transition function is not defined.

An \( \omega \text{OPBA} \) that recognizes \( L_1 \cap L_2 \) is defined in a similar way as for classical finite-state Büchi automata; precisely, \( A = (\Sigma, M = M_1 \cap M_2, Q, I, F, \delta) \), where
- \( Q = Q_1 \times Q_2 \times \{0, 1, 2\} \);
- \( I = I_1 \times I_2 \times \{0\} \);
- \( F = Q_1 \times Q_2 \times \{2\} \);
the transition function \( \delta : Q \times (\Sigma \cup Q) \rightarrow \mathcal{P}(Q) \) is defined as follows, where \( p_1, q_1, p_2, q_2 \in Q, a \in \Sigma \):
- \( \delta_{\text{push}}((p_1, p_2, x), a) = \{ (r_1, r_2, y) \mid r_1 \in \delta_{1\text{push}}(p_1, a) \land r_2 \in \delta_{2\text{push}}(p_2, a) \} \),
- \( \delta_{\text{shift}}((p_1, p_2, x), a) = \{ (r_1, r_2, y) \mid r_1 \in \delta_{1\text{shift}}(p_1, a) \land r_2 \in \delta_{2\text{shift}}(p_2, a) \} \),
- \( \delta_{\text{pop}}((p_1, p_2, x), (q_1, q_2, z)) = \{ (r_1, r_2, y) \mid r_1 \in \delta_{1\text{pop}}(p_1, q_1) \land r_2 \in \delta_{2\text{pop}}(p_2, q_2) \} \),
and the third component of the states is computed as follows:
- if \( x = 0 \) and \( r_1 \in F_1 \) then \( y = 1 \),
- if \( x = 1 \) and \( r_2 \in F_2 \) then \( y = 2 \),
- if \( x = 2 \) then \( y = 0 \),
- \( y = x \) otherwise.

Reading an input string, the automaton \( \mathcal{A} \) simulates \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \), respectively, on the first two components of the states, whereas the third component keeps track of the succession of visits of the two automata to their final states: in particular its value is 0 at the beginning, then switches from 0 to 1, from 1 to 2, and then back to 0, whenever the first automaton reaches a final state and the other one visits a final state afterwards. This cycle is repeated infinitely often whenever both the automata reach their final states infinitely many times along their run.

Conversely, if an \( \omega \)-word \( x \) does not belong to \( L_1 \cap L_2 \), then at least one of the runs of \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) must either stop because the transition function of the automaton is undefined for the given input or it does not visit infinitely often final states. Hence, \( \mathcal{A} \) cannot have a successful run on \( x \) and the word is rejected by \( \mathcal{A} \) too.

**Theorem 7.6** (\( \mathcal{L}(\omega \text{OPBA}) \) is closed under union). Let \( L_1 \) and \( L_2 \) be \( \omega \)-languages recognized by two \( \omega \text{OPBA} \) defined over the same alphabet \( \Sigma \), with compatible precedence matrices \( M_1 \) and \( M_2 \) and with \( s_1 \) and \( s_2 \) states, respectively. Then \( L = L_1 \cup L_2 \) is recognizable by an \( \omega \text{OPBA} \) with OPM \( M = M_1 \cup M_2 \) and \( O(\|\Sigma\|^2(s_1 + s_2)) \) states.

**Proof.** Let \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) be \( \omega \text{OPBA} \)s accepting \( L_1 \) and \( L_2 \) over OPMs \( M_1 \) and \( M_2 \), respectively. Without loss of generality we may assume \( M = M_1 = M_2 \) (otherwise one can apply Statement 7.1 increasing the number of states by a factor \( \|\Sigma\|^2 \)). For \( i = 1, 2 \), let \( \mathcal{A}_i = (\Sigma, M, Q_i, F_i, \delta_i) \). Then the \( \omega \)-language \( L = L_1 \cup L_2 \) is recognized by the \( \omega \text{OPBA} \) \( \mathcal{A} = (\Sigma, M, Q = Q_1 \cup Q_2, I = I_1 \cup I_2, F = F_1 \cup F_2, \delta) \) whose transition function \( \delta : Q \times (\Sigma \cup Q) \rightarrow \mathcal{P}(Q) \) is the nondeterministic union of \( \delta_1 \) and \( \delta_2 \), defined by setting \( \forall p, q \in Q, a \in \Sigma, \)

\[
\delta_{\text{push}}(q, a) = \begin{cases} \delta_{1\text{push}}(q, a) & \text{if } q \in Q_1, \\ \delta_{2\text{push}}(q, a) & \text{if } q \in Q_2, \end{cases} \\
\delta_{\text{shift}}(q, a) = \begin{cases} \delta_{1\text{shift}}(q, a) & \text{if } q \in Q_1, \\ \delta_{2\text{shift}}(q, a) & \text{if } q \in Q_2, \end{cases} \\
\delta_{\text{pop}}(p, q) = \begin{cases} \delta_{1\text{pop}}(p, q) & \text{if } p, q \in Q_1, \\ \delta_{2\text{pop}}(p, q) & \text{if } p, q \in Q_2, \end{cases}
\]

The above definition is well-posed since it applies to automata that share the same precedence matrix because they perform the same type of move (push/shift/pop) while reading the input word.

Since the sets of states of the two automata are disjoint and \( Q \) is their union, then for every \( x \in \Sigma^\omega \) there exists a successful run in \( \mathcal{A} \) iff there exists a successful run of \( \mathcal{A}_1 \) on \( x \) or a successful run of \( \mathcal{A}_2 \) on \( x \).

Clearly, the number of states of \( \mathcal{A} \) is \( |Q| = |Q_1| + |Q_2| \) and this concludes the proof, recalling the possible factor \( \|\Sigma\|^2 \) implied by Statement 7.1.

**Theorem 7.7** (closure of \( \mathcal{L}(\omega \text{OPBA}) \) under complementation). Let \( M \) be a conflict-free precedence matrix on an alphabet \( \Sigma \). Let \( L \) be an \( \omega \)-language on \( \Sigma \) that is recognized by a nondeterministic \( \omega \text{OPBA} \) with precedence matrix \( M \) and \( s \) states.
Then the complement of \( L \) w.r.t. \( L_M \) (the language of all the words \( x \in \Sigma^* \) compatible with \( M \)) is recognized by an \( \omega \)OPBA with the same precedence matrix \( M \) and \( 2O(s^2 + \Sigma|s \log 2|s) \) states.

Proof. The proof follows to some extent the structure of the corresponding proof for \( L(\omegaBVPA) \) [4], but it exhibits some relevant technical aspects which distinctly characterize it; in particular, we need to introduce an ad-hoc factorization of \( \omega \)-words due to the more complex management of the stack performed by OPAs.

Let \( A = (\Sigma, M, Q, I, F, \delta) \) be a nondeterministic \( \omega \)OPBA with \( |Q| = s \). Without loss of generality \( A \) can be considered complete w.r.t. the transition function \( \delta \), i.e., there is a run of \( A \) on every \( \omega \)-word on \( \Sigma \) compatible with \( M \).

An \( \omega \)-word \( w \in \Sigma^* \) compatible with \( M \) can be factored as a sequence of chains and pending letters \( w = w_1w_2w_3\ldots \), where either \( w_i = a_i \in \Sigma \) is a pending letter or \( w_i = a_1a_2\ldots a_{\text{in}} \) is the body of the chain \( i_1[w_i]_{\text{first}_i+1} \), where \( i_1 \) denotes the last pending letter preceding \( w_i \) in the word and \( \text{first}_i+1 \) denotes the first letter of word \( w_{i+1} \). Let also, by convention, \( a_0 = \# \) be the first pending letter.

Such factorization is not unique, since a string \( w_i \) can be nested into a larger chain having the same preceding pending letter. The factorization is unique, however, if we additionally require that the body \( w_i \) has no prefix (including itself) \( u_ib \) such that \( i_1[u_i]^b \) is a chain; in fact, in this case, as soon as a chain body with its context is identified after a pending letter, it becomes part of the factorization and what follows is either the beginning of a new body or a new pending letter.

For instance, for the word \( w = \langle a < c \rangle \langle a > \langle b > \rangle b \ldots \), with precedence relations in the OPM \( a \succ b \) and \( b < d \), two possible factorizations are \( w = w_1bw_2b \ldots \) and \( w = \overline{w}_1b\overline{w}_2\overline{w}_3b \ldots \), where \( b \) is a pending letter and \( \#[w_1]^b = \#[\overline{w}_1]^b = \#[ac]^b \), \( b_1[w_3]^3 = b_1[\overline{w}_3]^3 = b_1[a]^d \), and \( b_3[\overline{w}_4]^3 = b_3[d]^6 \) are chains. The second factorization is the unique one where each word \( \overline{w}_i \) has no prefix \( u_ib \) such that \( i_1[u_i]^b \) is a chain.

Let \( x \in \Sigma^* \) be the body of some chain \( a^x[\#]^b \) and let \( T(x) \) be the set of all triples \((q,p,f) \in Q \times Q \times \{0,1\} \) such that there exists a support \( q \xrightarrow{x} p \) in \( A \), and \( f = 1 \) iff the support contains a state in \( F \). Also let \( T \) be the set of all such \( T(x) \), i.e., \( T \) contains sets of triples identifying all supports for some chain, and set \( PR \) to be the finite alphabet \( \Sigma \cup T \). A pseudorum for the word \( w \) in \( A \)’s uniquely factorized as \( w_1w_2w_3 \ldots \) as stated above, is the \( \omega \)-word \( w' = y_1y_2y_3 \ldots \in PR^\omega \), where \( y_i = a_i \) if \( w_i \) is a pending letter, otherwise \( y_i = T(w_i) \).

For the unique factorization in the example above, then, \( w' = T(ac) b T(a) T(d) b \ldots \)

The automaton recognizing the complement of \( L = L(A) \) w.r.t. \( L_M \) can be built as an “online composition” of a transducer \( \omega \)OPBA \( B \) that computes the pseudorum corresponding to an input word \( w \), and a Büchi finite-state automaton \( B_R \) that recognizes all the pseudorums of \( \omega \)-words not accepted by \( A \): while reading \( w \), \( B \) outputs the pseudorum \( w' \) of \( w \) online, and the states of \( B_R \) are updated accordingly. The automaton accepts if both \( B \) and \( B_R \) reach infinitely often final states.

In order to define \( B_R \) we first define a nondeterministic Büchi finite-state automaton \( A_R = (PR, Q_{AR}, I_{AR}, F_{AR}, \delta_{AR}) \) over the alphabet \( PR \) whose language includes all pseudorums \( w' \) of any words \( w \in L(A) \).

The states of \( A_R \) correspond to the states of \( A \), but are extended with a lookback symbol and, in a correct pseudorum, represents the last pending letter of the input word read so far. \( A_R \) has all transitions corresponding to \( A \)’s push and shift transitions but is devoid of pop edges (in fact it is a finite-state automaton). In addition, for every \( S \in T \) it is endowed with arcs labeled \( S \) which link, for each triple \((q,p,f) \in S \)}
and \( a \in \hat{\Sigma} = \Sigma \cup \{\#\} \), either the pair of states \( \langle a, q \rangle, \langle a, p \rangle \) if \( f = 0 \), or \( \langle a, q \rangle, \langle a, p' \rangle \) if \( f = 1 \), where \( \langle a, p' \rangle \) is a new final state which takes into account the states in \( F \) met along the support \( q \rightarrow p \) and which has the same outgoing edges as \( \langle a, p \rangle \).

Formally, \( Q_{A_R} = \hat{\Sigma} \times (Q \cup Q') \), where \( Q' = \{q' \mid q \in Q\} \), \( I_{A_R} = \{\#\} \times I \), \( F_{A_R} = \hat{\Sigma} \times (F \cup Q') \). The transition function of \( A_R \) is defined as follows, where \( a \in \hat{\Sigma}, q \in Q, q' \in Q', S \in \mathcal{T} \) (\( \delta_{\text{push}} \) and \( \delta_{\text{shift}} \) are the transition functions of \( A \)):

- \( \delta(\langle a, q \rangle, b) = \begin{cases} \langle b, \delta_{\text{push}}(q, b) \rangle & \text{if } a < b, \\ \langle b, \delta_{\text{shift}}(q, b) \rangle & \text{if } a \geq b, \end{cases} \)
- \( \delta(\langle a, q \rangle, S) = \{\langle a, p \rangle \mid \langle q, p, 0 \rangle \in S\} \cup \{\langle a, p' \rangle \mid \langle q, p, 1 \rangle \in S\} \)
- \( \delta(\langle a, q' \rangle, X) = \delta(\langle a, q \rangle, X), \forall X \in PR \).

Notice that, given a set \( S \in \mathcal{T} \), the existence of an edge \( S \) between the pairs of states \( q, p \) in the triples in \( S \) can be decided in an effective way.

The automaton \( A_R \) built so far is able to parse all pseudoruns and recognizes all pseudoruns of \( \omega \)-words recognized by \( A \). However, since its moves are no longer completely determined by the OPM \( M \), it can also accept input words along the edges of the graph of \( A \) that are not pseudoruns since they do not correspond to a correct factorization on \( PR \). This is irrelevant, however, since the aim of the proof is to devise an automaton recognizing the complement of \( L(A) \), and all the words in \( L_M \setminus L(A) \) are parsed along pseudoruns, which are not accepted by \( A_R \). If one gives as input words only pseudoruns (and not generic words on \( PR \)), then they will be accepted by \( A_R \) if the corresponding words on \( \Sigma \) belong to \( L(A) \), and they will be rejected if the corresponding words do not belong to \( L(A) \) (see Figure 20). Then we can construct a deterministic \( \text{B"uchi} \) automaton \( B_R \) that accepts the complement of \( L(A_R) \) on the alphabet \( PR \) [36]. If \( B_R \) receives only input words on \( PR \) that are pseudoruns, then it will accept only words in \( L_M \setminus L(A) \).

Now we define a nondeterministic transducer \( \omega \text{OPBA} \) \( B \) which on reading \( w \) generates online the pseudorun \( w' \). The transducer \( B \) nondeterministically guesses whether the next input symbol is a pending letter, the beginning of a chain appearing in the factorization of \( w \), or a symbol within such a chain, and uses stack symbols \( Z \), \( B \), or elements in \( \mathcal{T} \), respectively, to distinguish these three cases.

Whenever the automaton reads a pending letter it outputs the same letter, whereas when it completes the recognition of a chain of the factorization, performing a pop move that removes from the stack an element with state \( B \), it outputs the set of all the pairs of states which define a support for the chain. Thus, the output \( w' \) produced by \( B \) is unique, despite the nondeterminism of the transducer.
Formally, the transducer $\omega OPBA \mathcal{B} = \langle \Sigma, M, Q_B, I_B, F_B, PR, \delta_B, \eta_B \rangle$ is defined as follows:

- $Q_B = \{ Z, B \} \cup \mathcal{T}$, i.e., a state in $Q_B$ represents the guess of whether the next symbol to be read is a pending letter ($Z$), the beginning of a chain ($B$), or a letter within such a chain $w_i$ ($T \in \mathcal{T}$). In the third case, $T$ contains all information necessary to correctly simulate the moves of $A$ during the parsing of the chain $w_i$ of $w$, and compute the corresponding symbol $y_i$ of $w'$. In particular, $T$ is a set comprising all triples $(r, q, \nu)$, where $r$ represents the state reached before the last push move, $q$ represents the current state reached by $A$, and $\nu$ is a bit that reminds us of whether, while reading the chain, a state in $F$ has been encountered (as in the construction of a deterministic OPA on words of finite length, it is necessary to keep track of the state from which the parsing of a chain started, to avoid erroneous merges of runs on pop moves).

- $I_B = F_B = \{ B, Z \}$.

- The transition function and the output function are defined as the union of three pairs of functions. Let $\langle \delta_B^{push}, \eta_B^{push} \rangle: Q_B \times \Sigma \rightarrow \mathcal{W}(Q_B \times PR^*)$ is defined as follows, where the symbols after $\downarrow$ denote the output.

  - Push of a pending letter.
    
    \[ \langle \delta_B^{push}, \eta_B^{push} \rangle(Z, a) = \{ B \downarrow a, Z \downarrow a \} \, . \]

  - Push at the beginning of a chain of the factorization.
    
    \[ \langle \delta_B^{push}, \eta_B^{push} \rangle(B, a) = \{ T \downarrow \varepsilon \} \, , \]
    
    where $T = \{ (q, p, \nu) \mid q \in Q, p \in \delta_{push}(q, a), \nu = 1 \text{ iff } p \in F \}$.

  - Push within a chain of the factorization.
    
    \[ \langle \delta_B^{push}, \eta_B^{push} \rangle(T, a) = \{ S \downarrow \varepsilon \} \, , \quad \text{where} \]
    
    \[ S = \left\{ (q, p, \nu) \mid \exists (r, q, \xi) \in T \text{ s.t. } \nu = \begin{cases} \xi & \text{if } p \notin F, \\ 1 & \text{if } p \in F, \end{cases} \right\} \, . \]

The shift pair $\langle \delta_B^{shift}, \eta_B^{shift} \rangle: Q_B \times \Sigma \rightarrow \mathcal{W}(Q_B \times PR^*)$ is defined as follows.

- Pending letter.
    
    \[ \langle \delta_B^{shift}, \eta_B^{shift} \rangle(Z, a) = \{ B \downarrow a, Z \downarrow a \} \, . \]

- Shift move within a chain of the factorization.
    
    \[ \langle \delta_B^{shift}, \eta_B^{shift} \rangle(T, a) = \{ S \downarrow \varepsilon \} \, , \quad \text{where} \]
    
    \[ S = \left\{ (r, p, \nu) \mid \exists (r, q, \xi) \in T \text{ s.t. } \nu = \begin{cases} \xi & \text{if } p \notin F, \\ 1 & \text{if } p \in F, \end{cases} \right\} \, . \]

The pop pair $\langle \delta_B^{pop}, \eta_B^{pop} \rangle: Q_B \times Q_B \rightarrow \mathcal{W}(Q_B \times PR^*)$ is defined as follows.

- Pop at the end of a chain of the factorization.
    
    \[ \langle \delta_B^{pop}, \eta_B^{pop} \rangle(T, B) = \{ B \downarrow R, Z \downarrow R \} \, , \quad \text{where} \]
    
    \[ R = \left\{ (r, p, \nu) \mid \exists (r, q, \xi) \in T \text{ s.t. } p \in \delta_{pop}(q, r), \nu = \begin{cases} \xi & \text{if } p \notin F, \\ 1 & \text{if } p \in F \end{cases} \right\} \, . \]
- **Pop within a chain of the factorization.**

\[
\langle \delta_{\text{B}_{\text{pop}}}, \eta_{\text{B}_{\text{pop}}} \rangle(T, S) = \{ R \downarrow \varepsilon \}, \quad \text{where}
\]

\[
R = \left\{ (t, p, \nu) \mid \exists (r, q, \xi) \in T, \exists (t, r, \zeta) \in S \text{ s.t. } p \in \delta_{\text{pop}}(q, r), \right. \]

\[
\left. \nu = \begin{cases} 
\xi & \text{if } p \not\in F \\
1 & \text{if } p \in F 
\end{cases} \right\},
\]

An error state is reached in any other case.

We conclude the construction by computing the size of the resulting automaton, which is an “online composition” of \( B \) and \( B_R \). The Büchi finite-state automaton \( A_R \) has \( O(|\Sigma|^s) \) states and hence the automaton \( B_R \) has \( 2^{O(|\Sigma|^s \log |\Sigma|^s)} \) states \([40, 36]\), whereas the transducer \( B \) has \( |Q_B| = 2^{O(s^2)} \) states. Thus the \( \omega \text{OPBA} \) has \( 2^{O(s^2 + |\Sigma|^s \log |\Sigma|^s)} \) states.

To prove that \( B \) produces all \( A \)'s pseudoruns—whether accepting or not—observe, first that its guess about reading a pending letter, or the beginning of a chain belonging to the unique factorization defined above, or reading a symbol within such a chain, is essentially the same as the one described in the proof of Lemma 7.3, where the recognition of a maximal chain is replaced by the recognition of a chain with no prefixes that are chains; thus, wrong guesses are resolved at the time of a pop move (e.g., a pop move is not defined on a first state of type \( Z \)). Furthermore, pending letters, when correctly guessed as such, are output as soon as they are read (the incorrectly guessed ones belong to runs that will be aborted); elements of \( T \) are output only when a chain of the factorization is recognized, i.e., the transition is defined on a pair of states whose second component is \( B \), which separates these moves from the pop ones occurring within a chain of the factorization; the set \( T \) output during the move records all pairs of states that can be the beginning and the end of a support of the recognized chain. Finally the input string is accepted if infinitely many times either pending letters are read or chains of the factorization are recognized, or both facts occur, i.e., the string is compatible with the OPM, and the produced output is the pseudorun associated with the input by definition, independently of whether the original \( A \)'s run was accepting, i.e., infinitely many times sets of triples with \( \nu = 1 \) have been output, or not.

Let us finally consider the case of concatenation between a finite-length OPL and a language in \( \mathcal{L}(\omega \text{OPBA}) \). For classical families of automata (on finite- or infinite-length words) the closure with respect to concatenation is traditionally proved by building an automaton which simulates the moves of the first automaton while reading the first word of the concatenation and—whether deterministically or not—once it reaches some final state, it switches to the initial states of the second one. This natural approach has already been proved ineffective for OPLs in the case of finite-length words since the structure of two concatenated strings is not necessarily the concatenation of the two structures, so that the actions of the second automaton cannot be independent from those of the previous one (\([18]\) provides a constructive proof of the closure of finite-length OPLs w.r.t. concatenation in terms of generating grammars); in fact, the lack of the \( \# \) delimiter between the two strings prevents the typical lookahead mechanisms which drive the operator-based parsing; thus, the stack cannot be emptied by the normal sequence of pop moves before beginning the parse.

---

10Remember that we consider only chains having no prefixes that are chains.
of the new string. In the case of \( \omega \)-languages the difficulty is further exacerbated by the fact the automaton might never be able to empty the stack as, e.g., in the case of a language \( L_1 \subseteq \{a, b\}^* \) with \( a \prec a, b \prec a \), concatenated with \( L_2 = \{a^\omega\} \). Notice also that, after reading the first finite word in the concatenation, it would not be possible to determine whether this word might be accepted by—possibly nondeterministically—guesing the position of a potential delimiter \( \# \), since this check would require us to know the states already reached and piled on the stack, which are not visible without emptying the stack itself.

To overcome the above difficulties we follow this approach:

- We give up deterministic parsing. In fact the different computational power between deterministic and nondeterministic automata is a distinguishing property when moving from finite- to infinite-length languages. Thus, we nondeterministically guess the point of separation between the first finite word and the second infinite one.
- To afford the second major problem, i.e., the lack of enough knowledge to decide whether the guessed first word would be accepted by the corresponding automaton, we use \( \# \)OPAs introduced in section 7.1.

The following theorem exploits the above approach. Its proof differs significantly from the nontrivial proof of closure under concatenation of OPLs of finite-length words [18], which, instead, can be recognized deterministically.

**Theorem 7.8** (\( L(\omega \text{OPBA}) \) is closed under concatenation). Let \( L_1 \subseteq \Sigma^* \) be a language of finite words recognized by an OPA with OPM \( M_1 \) and \( s_1 \) states. Let \( L_2 \subseteq \Sigma^* \) be an \( \omega \)-language recognized by a nondeterministic \( \omega \)OPBA with OPM \( M_2 \) compatible with \( M_1 \) and \( s_2 \) states. Then the concatenation \( L_1 \cdot L_2 \) is also recognized by an \( \omega \)OPBA with OPM \( M \supseteq M_1 \cup M_2 \) and \( O(s_1^2 + s_2^2) \) states.

**Proof.** Let \( A_1 \) be a nondeterministic OPA on \((\Sigma, M_1)\) that recognizes \( L_1 \) and let \( A_2 = (\Sigma, M_2, Q_2, I_2, F_2, \delta_2) \) be a nondeterministic \( \omega \)OPBA with OPM \( M_2 \) compatible with \( M_1 \) that accepts \( L_2 \). Suppose, without loss of generality, that the sets of states of \( A_1 \) and \( A_2 \) are disjoint.

To define an automaton \( \omega \)OPBA \( A \) that accepts \( L_1 \cdot L_2 \), we first build a \( \# \)OPA \( A'_1 = (\Sigma, M_1, Q_1, I_1, F_1, \delta_1) \) such that \( L_{\#}(A'_1) = L(A_1) \).

The automaton \( A \) can recognize the first finite words in the concatenation \( L_1 \cdot L_2 \) by simulating \( A'_1 \): reading the input string, if \( A'_1 \) reaches a final state at the end of a finite-length prefix, then it belongs to \( L_1 \) and \( A \) immediately starts the recognition of the second infinite string without the need to perform any pop move to empty the stack. From this point onwards, then, \( A \) checks that the remaining infinite portion of the input belongs to \( L_2 \), behaving as the \( \omega \)OPBA \( A_2 \).

The strings belonging to the concatenation of two OPLs, however, may contain new chains that span over the two concatenated words. Consider, for instance, the concatenation of \( L_1 = \{a^m b^n \mid m \geq n \geq 1\} \) with \( L_2 = \{c^i b^\omega\} \); notice that any OPA recognizing \( L_1 \) must be defined on an OPM such that \( a \prec a, a \doteq b, b \succ b \) to be able to compare the occurrences of \( a \) with those of \( b \); assume also the further precedence relations \( a \prec c, c \prec c, c \succ b \) (such relations could be mandated, e.g., by other components of either language not included here for simplicity). An automaton recognizing \( L_1 \cdot L_2 \) can deterministically find the borderline between words \( x \in L_1 \), and \( y \in L_2 \); after finishing reading \( x \) it will have on its stack \( m - n \) remaining \( a \)'s; however, since \( a \prec c \) it cannot empty the stack and must push all \( c \)'s on top of the \( a \)'s. Only when receiving the first \( b \), it will pop all \( c \)'s until the top of the stack will store an \( a \). Since \( a \doteq b \), and \( b \succ b \) the next action must consist in shifting the \( b \) by replacing the topmost \( a \) and then popping it, thus consuming part of the stack left by the analysis of \( x \); in
other words, it must produce the support of a chain \(a[ab]^b\), whose left part belongs to \(L_1\) and whose right part belongs to \(L_2\).

Therefore, \(A\) cannot merely read the second infinite word performing the same transitions as \(A_2\), but it can still simulate this \(\omega\)OPBA by keeping in the states some summary information about its runs. In this way, while reading the second word in the concatenation, whenever \(A\) has to reduce a chain that extends to the previous word in \(L_1\) and, therefore, must perform a pop move of a symbol in the portion of the stack piled up during the parsing of the first word, it can restore the state that \(A_2\) would instead have reached, resuming therefrom as in a run of \(A_2\).

Precisely, \(A\) is defined as the tuple \((\Sigma, M, Q, I, F, \delta)\) where we have the following:

- \(M \supseteq M_1 \cup M_2\) and may be supposed to be a complete matrix, for instance, assigning arbitrary precedence relations to the empty entries, so that the strings in the concatenation of languages \(L_1\) and \(L_2\) are compatible with \(M\).
- \(Q = Q_1 \cup Q_2 \cup Q_2 \times (Q_2 \cup \{-\})\), i.e., the set of states of \(A\) includes the states of \(A_1\) and \(A_2\), along with the states of \(A_2\) extended with a second component. The first component is the state of \(Q_2\) that \(A_2\) would reach in its corresponding computation on the second word of the concatenation, and the second one represents the state of the symbol that is on the top of the stack when the current input letter is read in this run of \(A_2\). Storing this component is necessary to guarantee that, whenever the automaton \(A\) has to perform a pop move that removes symbols that have been piled on the stack during the recognition of the first word in the concatenation, it is still possible to compute the state that \(A_2\) would have reached instead.

This second component is denoted ‘‘\(0\)’’ if all the preceding symbols in the stack have been piled up during the parsing of the first word of the concatenation (thus the stack of \(A_2\) is empty).

- \(I = I_1 \cup \{\langle q_0, - \rangle \mid q_0 \in I_2\} \text{ if } \varepsilon \in L_1; I = I_1 \text{ otherwise.}\)
- \(F = F_2 \cup F_2 \times (Q_2 \cup \{-\})\).

The transition function \(\delta : Q \times (\Sigma \cup Q) \to \mathcal{P}(Q)\) is \(\delta = \delta_1 \cup \delta_2 \cup \delta_{\text{join}}\) and is defined as the union of three functions: the transition functions of \(A_1\) and \(A_2\) by which it simulates the first automaton on the first word of the concatenation and the second automaton on the second one, and a function \(\delta_{\text{join}}\) that handles the nondeterministic transition from the simulation of the first automaton to the second one and the parsing of the suffix (within the second word of the concatenation) of the chains that span over the two words.

**Function \(\delta_{\text{join}}\) is defined as follows:** let \(c \in \Sigma, p \in Q_1, q_1, q_2, q_3 \in Q_2, r \in (Q_2 \cup \{-\})\).

The push transition function \(\delta_{\text{push}} : Q \times \Sigma \to \mathcal{P}(Q)\) is defined by

- \(\delta_{\text{push}}(p, c) = \{\langle q_0, - \rangle \mid q_0 \in I_2\} \text{ if } \exists p_f \in F_1 \text{ s.t. } \delta_{\text{push}}(p, c) \ni p_f\),
- \(\delta_{\text{push}}(q, r, c) = \delta_{2 \text{push}}(q, c)\),

i.e., \(A\) nondeterministically enters the initial states of \(A_2\) after the recognition of a word in \(L_1\):

The shift transition function \(\delta_{\text{shift}} : Q \times \Sigma \to \mathcal{P}(Q)\) is defined by

- \(\delta_{\text{shift}}(p, c) = \{\langle q_0, - \rangle \mid q_0 \in I_2\} \text{ if } \exists p_f \in F_1 \text{ s.t. } \delta_{\text{shift}}(p, c) \ni p_f\),

i.e., \(A\) nondeterministically enters the initial states of \(A_2\) after the recognition of a word in \(L_1\):
Table 8
Size of state sets of languages recognizing \( L_1 \cap L_2 \), \( L_1 \cup L_2 \), \( \neg L_1 \), and \( L_3 \cdot L_1 \). The results on \( \omega \)OPBAs have been proved, respectively, in Theorems 7.5, 7.6, 7.7, and 7.8. The complexity results on \( \omega \)BVPAs derive from the constructions and proofs of their closure properties shown in [4].

<table>
<thead>
<tr>
<th>( L_1 \cap L_2 )</th>
<th>( L_{\omega \text{OPBA}} )</th>
<th>( L_{\omega \text{BPA}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_1 \cup L_2 )</td>
<td>( O(s_1 s_2) )</td>
<td>( O(s_1 s_2) )</td>
</tr>
<tr>
<td>( \neg L_1 )</td>
<td>( 2^O(s_1^2 +</td>
<td>\Sigma</td>
</tr>
<tr>
<td>( L_3 \cdot L_1 )</td>
<td>( O(s_1^2 + s_2^2) )</td>
<td>( s_1 + s_3 )</td>
</tr>
</tbody>
</table>

- \( \delta_{\text{join}}^\omega(\langle q_1, - \rangle, c) = \{ \langle q_2, q_1 \rangle \mid q_2 \in \delta_{\text{push}}(q_1, c) \} \), i.e., \( A \) simulates the push move induced by the precedence relation \( \# c \) that, in the corresponding run of \( A_2 \), starts the recognition of a chain that is a prefix of the second word of the concatenation;

- \( \delta_{\text{shift}}^\omega(\langle q_1, q_2 \rangle, c) = \{ \langle q_3, q_2 \rangle \mid q_3 \in \delta_{\text{shift}}(q_1, c) \} \), i.e., \( A \) performs a shift move within a chain that spans over the two words of the concatenation.

The pop transition function \( \delta_{\text{pop}}^\omega : Q \times Q \to \wp(Q) \) is defined by

- \( \delta_{\text{pop}}^\omega(\langle q_1, - \rangle, p) = \langle q, - \rangle \), i.e., \( A \) concludes to recognize a chain, at the end of the first word of the concatenation, induced by the precedence relations with the letters of the second string, and consumes the corresponding stack symbols piled while reading the first word;

- \( \delta_{\text{pop}}^\omega(\langle q_1, q_2 \rangle, p) = \{ \langle q_3, - \rangle \mid q_3 \in \delta_{\text{pop}}(q_1, q_2) \} \), i.e., whenever the precedence relations induce a merge of the chains of the words of the concatenation, \( A \) restores the state \( q_3 \) of \( A_2 \) from which a run of \( A_2 \) will continue;

- \( \delta_{\text{join}}^\omega(\langle q_1, q_2 \rangle, r) = \{ \langle q_3, r \rangle \mid q_3 \in \delta_{\text{join}}(q_1, q_2) \} \), i.e., \( A \) completes the recognition of a chain that belongs to a composed chain spanning over the two words of the concatenation.

One can verify that, after having simulated \( A_1 \) and nondeterministically guessed the end of a word in \( L_1 \), \( A \) proceeds with the simulation of \( A_2 \) and accepts the remaining \( \omega \)-string if it belongs to \( L_2 \). In fact, the projection on the first component of the states visited along \( A \)'s run on the second word of the concatenation identifies a successful run of \( A_2 \) on the same word.

To summarize, Table 8 displays the complexities of the various constructions to obtain the closure w.r.t. Boolean operations and concatenation; it also compares them with the corresponding complexities for VPLs showing that the only main difference occurs in the case of concatenation.

### 7.3. Closure properties of the other classes of \( \omega \)OPLs.

The class of languages recognized by \( \omega \)DOPMAs is a Boolean algebra. The other classes are closed only under union and intersection.

**Theorem 7.9** (\( L_{\omega \text{DOPMA}} \) is a Boolean algebra). Let \( L_1 \) and \( L_2 \) be \( \omega \)-languages that are recognized by two \( \omega \)DOPMAs defined over the same alphabet \( \Sigma \), with compatible precedence matrices \( M_1 \) and \( M_2 \) and \( s_1 \) and \( s_2 \) states, respectively. Then \( L_1 \cap L_2 \) (resp., the complement of \( L_1 \) w.r.t. \( L_M \), or \( L_1 \cup L_2 \)) is recognized by an \( \omega \)DOPMA with \( OPM = M_1 \cap M_2 \) and \( s_1 s_2 \) (resp., \( s_1 \), or \( |\Sigma|^4 s_1 s_2 \)) states.
Proof. Let \( A_1 = (\Sigma, M_1, Q_1, q_{01}, T_1, \delta_1) \) and \( A_2 = (\Sigma, M_2, Q_2, q_{02}, T_2, \delta_2) \) be \( \omega \)-DOPMAs recognizing languages \( L_1 \) and \( L_2 \). Assume without loss of generality that their transition function is total (otherwise, it can be naturally completed once the set of states is extended with an “error” state).

An \( \omega \)-DOPMA \( A \) with OPM \( M = M_1 \cap M_2 \) recognizing \( L = L_1 \cap L_2 \) may be defined adopting the usual product construction for \( \omega \)-regular automata, requiring that a successful path in \( A \) corresponds to paths that visit infinitely often sets in the table \( T_1 \) and \( T_2 \). More precisely let \( \mathcal{A} = (\Sigma, M, Q, q_0, T, \delta) \), where

- \( Q = Q_1 \times Q_2 \);
- \( q_0 = (q_{01}, q_{02}) \);
- \( \pi_i \) (\( i = 1, 2 \)) is defined as the projection from \( Q_1 \times Q_2 \) on \( Q_i \), that can also be naturally extended to define projections on paths of the automata, and let \( T = \{ P \subseteq Q_1 \times Q_2 \mid \pi_1(P) \in T_1 \land \pi_2(P) \in T_2 \} \);
- the transition function \( \delta \) is the product of \( \delta_1 \) and \( \delta_2 \) (see Definition 7.1).

Let \( \rho \) be a successful path of \( \mathcal{A} \), starting in the initial state \( q_0 = (q_{01}, q_{02}) \); since it is accepting, the set \( Inf(\rho) = P \in T \). By the definition of \( T \), the paths \( \rho_1 \) and \( \rho_2 \) that are the projection of \( \rho \) on the set of states of \( A_1 \) and \( A_2 \), respectively, have \( Inf(\rho_1) = \pi_1(P) \in T_1 \) and \( Inf(\rho_2) = \pi_2(P) \in T_2 \); hence \( \rho_1 \) and \( \rho_2 \) are successful paths for the two automata, and \( x \) belongs to \( L(A_1) \cap L(A_2) \).

Let now \( x \in L(A_1) \cap L(A_2) \); thus, \( x \) labels two successful paths \( \rho_1 \) and \( \rho_2 \) of the two automata, i.e., \( Inf(\rho_1) \in T_1 \) and \( Inf(\rho_2) \in T_2 \). The path \( \rho \) of \( \mathcal{A} \) which visits the pairs of states of the two automata, performing the same type of move they perform for each input symbol, is defined as \( \pi_1(Inf(\rho)) = \pi_1(\rho_1) \in T_1 \) and \( \pi_2(Inf(\rho)) = \pi_2(\rho_2) \in T_2 \). Therefore, by the definition of \( T \), \( \rho \) is a successful path for \( \mathcal{A} \).

To recognize the complement of \( L_1 \), given that \( \mathcal{A} \) is deterministic and its transition function is total, it is clearly sufficient to build the \( \omega \)-DOPMA \( \mathcal{A}' = (\Sigma, M_1, Q_1, q_{01}, O(Q_1) \setminus T_1, \delta) \) whose table is the complement of \( T_1 \) w.r.t. \( O(Q_1) \).

To obtain the closure w.r.t. union, we can assume that \( M_1 = M_2 \) w.l.o.g. (otherwise one can apply Statement 7.1, increasing the number of states of each automaton of a factor \( |\Sigma|^2 \)) and apply De Morgan’s law. The number of states of the resulting automaton is \( |Q_1| \cdot |Q_2| \) and this concludes the proof, recalling the factor \( |\Sigma|^2 \cdot |\Sigma|^2 \) implied by the possible application of Statement 7.1. Notice that, if one considers automata with compatible but not equal matrices, De Morgan’s law could not be applied; in fact, the equality

\[
L_1 \cup L_2 = L_{M_1 \cap M_2} \setminus [(L_{M_1} \setminus L_1) \cap (L_{M_2} \setminus L_2)]
\]

does not hold, unless \( M_1 = M_2 \). \( \square \)

Proposition 7.10. Let \( L_1 \) and \( L_2 \) be \( \omega \)-languages recognized by two \( \omega \)OPBEA (resp., \( \omega \)DOPBA, \( \omega \)DOPBEA) defined over the same alphabet \( \Sigma \), with compatible precedence matrices \( M_1 \) and \( M_2 \) and with \( s_1 \) and \( s_2 \) states, respectively. Then \( L = L_1 \cap L_2 \) is recognized by an \( \omega \)OPBEA (resp., \( \omega \)DOPBA, \( \omega \)DOPBEA) with OPM \( M = M_1 \cap M_2 \) and \( O(s_1s_2) \) states.

Proof. For \( \omega \)OPBEA, we can assume without loss of generality that the automaton is in normal form with partitioned sets of states (see Definition 5.5), and apply the same construction as for \( \omega \)OPBA (see Theorem 7.5). The use of automata with partitioned sets of states guarantees that a run of \( \mathcal{A} \) on an \( \omega \)-word reaches infinitely often a final state with empty stack iff both \( A_1 \) and \( A_2 \) have a run for the word which traverses infinitely often a final state with empty stack.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
For \( \omega \text{DOPBA} \) and \( \omega \text{DOPBEA} \), the proof derives from the fact that, if \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) are deterministic, then the resulting intersection automaton is deterministic too.

**Proposition 7.11.** Let \( L_1 \) and \( L_2 \) be \( \omega \)-languages recognized by two \( \omega \text{OPBEA} \) (resp., \( \omega \text{DOPBA}, \omega \text{DOPBEA} \)) defined over the same alphabet \( \Sigma \), with compatible precedence matrices \( M_1 \) and \( M_2 \) and \( s_1 \) and \( s_2 \) states, respectively. Then \( L = L_1 \cup L_2 \) is recognized by an \( \omega \text{OPBEA} \) (resp., \( \omega \text{DOPBA}, \omega \text{DOPBEA} \)) with OPM \( M = M_1 \cup M_2 \) and \( O((\Sigma^2 s_1 s_2)) \) (resp., \( O((\Sigma^2 s_1 s_2)) \)) states.

**Proof.** The proof for \( \omega \text{OPBEA} \) is analogous to the proof of closure under union for \( \omega \text{OPBA} \) (see Theorem 7.5).

For the deterministic models, the construction must be refined. Let \( \mathcal{A}_1 \) and \( \mathcal{A}_2 \) be \( \omega \text{DOPBA} \) accepting \( L_1 \) and \( L_2 \) over OPMs \( M_1 \) and \( M_2 \), respectively. As usual, we assume that both transition functions are complete and \( M_1 = M_2 \) (otherwise one can apply Statement 7.1, increasing the number of states of a factor \( |\Sigma|^2 \)). Let \( \mathcal{A}_3 = (\Sigma, M, Q_3, q_{03}, F_3, \delta_3) \) for \( i = 1, 2 \). An \( \omega \text{DOPBA} \) (resp., \( \omega \text{DOPBEA} \)) \( \mathcal{A}_3 \) which recognizes \( L_1 \cup L_2 \) is then defined by adopting the usual product construction for regular automata, \( \mathcal{A}_3 = (\Sigma, M, Q_3, q_{03}, F_3, \delta_3) \), where

\[
Q_3 = Q_1 \times Q_2,
\]
\[
q_{03} = (q_{01}, q_{02}),
\]
\[
F_3 = F_1 \times Q_2 \cup Q_1 \times F_2,
\]
\[\text{and the transition function is the product of } \delta_1 \text{ and } \delta_2.\]

The number of states of \( \mathcal{A}_3 \) is given by the product \( |Q_1| \cdot |Q_2| \) and this concludes the proof, recalling the factor \( |\Sigma|^2 \cdot |\Sigma|^2 \) implied by Statement 7.1.

**Theorem 7.12 (\( \omega \text{DOPBA}, \omega \text{OPBEA}, \omega \text{DOPBEA} \) are not closed under complement).** Let \( L \) be an \( \omega \)-language accepted by an \( \omega \text{DOPBA} \) (resp., \( \omega \text{OPBEA} \) or \( \omega \text{DOPBEA} \)) with OPM \( M \) on alphabet \( \Sigma \). There does not necessarily exist an \( \omega \text{OPBEA} \) (resp., \( \omega \text{OPBA} \) or \( \omega \text{DOPBEA} \)) recognizing the complement of \( L \) w.r.t. \( L_M \).

**Proof.** Language \( L_{a\infty} \) can be recognized by an \( \omega \text{DOPBA} \) with OPM \( M \) (shown, for instance, in Figure 15), but there’s no \( \omega \text{DOPBA} \) that can recognize the complement of this language w.r.t. \( L_M \), i.e., the language \( L_{a\text{--finite}} \), as mentioned in section 6.2. The same argument on \( L_{a\infty} \) holds also for \( \omega \text{DOPBEAs} \).

Finally, as regards \( \omega \text{OPBEAs} \), \( L_{a\text{--seq}} \) is recognized by the \( \omega \text{OPBEA} \) with OPM \( M \) and state graph presented in section 6.2. However, no \( \omega \text{OPBEA} \) can recognize the complement of this language w.r.t. \( L_M \). Such an \( \omega \text{OPBEA} \), in fact, should have OPM \( M \) so that no word in \( L_{a\text{--seq}} \) can be accepted. The precedence relation \( M_{aa} = \{<\} \) (which is necessary to verify that in a sequence of type \( (a^k b^h)^n \) there is at least one substring with \( k \neq h \)), however, prevents an \( \omega \text{OPBEA} \) from accepting the word \( a^n \), which belongs to the complement of \( L_{a\text{--seq}} \) w.r.t. \( L_M \), since it implies that, while reading the word, the \( \omega \text{OPBEA} \) can never reach a state with empty stack.

**Theorem 7.13 (\( \omega \text{DOPBA}, \omega \text{OPBEA}, \omega \text{DOPBEA}, \text{and } \omega \text{DOPMA} \) are not closed under concatenation).** Let \( L_2 \) be an \( \omega \)-language accepted by an \( \omega \text{OPBEA} \) (resp., \( \omega \text{DOPBA}, \omega \text{DOPBEA}, \text{or } \omega \text{DOPMA} \)) with OPM \( M \) on alphabet \( \Sigma \) and let \( L_1 \subseteq \Sigma^* \) be a language (of finite words) recognized by an OPA with a compatible precedence matrix. The \( \omega \)-language defined by the concatenation \( L_1 \cdot L_2 \) is not necessarily recognizable by an \( \omega \text{OPBEA} \) (resp., \( \omega \text{DOPBA}, \omega \text{DOPBEA}, \text{or } \omega \text{DOPMA} \)).

**Proof.** For \( \omega \text{DOPBAs} \), let \( \Sigma = \{a, b\} \) and consider the language \( L_{a\text{--finite}} \), which can be seen as the concatenation \( L_{a\text{--finite}} = L_1 \cdot L_2 \) of a language of finite words \( L_1 = \{a, b\}^* \), which can be clearly recognized by an OPA, and an \( \omega \)-language \( L_2 = \{b^n\} \), which can be recognized by an \( \omega \text{DOPBA} \) with compatible precedence matrices. Since
language \( L_{\omega} \)-finite cannot be recognized by an \( \omega \)DOPBA, then the class of languages \( \mathcal{L}(\omega \)DOPBA) is not closed w.r.t. concatenation.

Given \( \Sigma = \{c,r\} \), the language \( L_{\text{repbsd}} \) cannot be recognized by an \( \omega \)OPBEA (resp., \( \omega \)DOPBEA or \( \omega \)DOPMA), as shown in section 6.2. Consider the OPA that accepts the language \( L_1 = \Sigma^* \) of words of finite length whose OPM is the same as the precedence matrix depicted in Figure 16. These words have necessarily a finite number of pending calls, since they have finite length. Moreover, let \( A_2 \) be an \( \omega \)OPBEA (resp., \( \omega \)DOPBEA or \( \omega \)DOPMA) that recognizes the \( \omega \)-language \( L_{\text{Dyck-pr}(c,r)} \) and which is depicted in Figure 16. The concatenation \( \omega \)-language \( L_1 \cdot L_{\text{Dyck-pr}(c,r)} \) is exactly the set of \( \omega \)-words with a finite number of pending calls, i.e., \( L_{\text{repbsd}} \). Hence, the class of languages \( \mathcal{L}(\omega \)OPBEA) (resp., \( \mathcal{L}(\omega \)DOPBEA) or \( \mathcal{L}(\omega \)DOPMA)) is not closed w.r.t. concatenation.

8. MSO logic characterization of \( \omega \)OPLs. We finally provide a characterization of \( \omega \)OPLs in terms of an MSO logic which is interpreted over infinite words. As usual, we focus our attention on \( \mathcal{L}(\omega \)OPBA), the most general class of \( \omega \)OPLs.

We adopt the same conventions and notation as in section 4, and extend the formula evaluation over \( \omega \)-strings in the natural way. To distinguish the infinite case from the finite one, we will use the symbol \( \models_{\omega} \) instead of \( \models \). Given an OP alphabet \( (\Sigma,M) \) and an MSO formula \( \varphi \), we denote the language of all strings \( w \in \Sigma^\omega \) such that \( \# w \models_{\omega} \varphi \) by \( L^\omega(\varphi) = \{ w \in \Sigma^\omega \mid \# w \models_{\omega} \varphi \} \).

Example 10 (managing interrupts). Consider again the system that manages interrupts described in Example 8. The same rules enforced by the automaton of Figure 13 are also formalized by the following sentences.

- All \( \text{int}_2 \) are eventually served by a corresponding \( \text{serve}_2 \):
  \[
  \forall x \ (\text{int}_2(x) \Rightarrow \exists y(\text{serve}_2(y) \land (y = x + 1 \lor x \vartriangleleft y))).
  \]

- Lower priority interrupts are not accepted when a higher priority one is pending:
  \[
  \forall x, y \ (\text{int}_2(x) \land \text{serve}_2(y) \land x \vartriangleleft y \Rightarrow \forall k (x < k < y \Rightarrow \neg \text{int}_1(k))).
  \]

As another example consider the “weak fairness requirement” also mentioned in Example 8 which states that after a first \( \text{call}_a \) not matched by \( \text{ret}_a \) but interrupted by an \( \text{int}_1 \) or \( \text{int}_2 \), a second \( \text{call}_a \) cannot be interrupted by a new lower priority interrupt \( \text{int}_1 \) (but can still be interrupted at any time by higher priority ones): the sentence below formalizes such a constraint.

\[
\neg \exists x_1, x_2 (x_1 < x_2 \land \text{call}_a(x_1) \land \text{call}_a(x_2) \\
\land \forall x_3 (x_1 \leq x_3 \leq x_2 \land \text{call}_a(x_3) \Rightarrow \neg \exists y_3 (\text{ret}_a(y_3) \land (y_3 = x_3 + 1 \lor x_3 \vartriangleleft y_3)) \\
\land \exists z_1, z_2 ((\text{int}_1(z_1) \lor \text{int}_2(z_1)) \land \text{int}_1(z_2) \land \bigwedge_{i=1}^{2} (z_i = x_i + 1 \lor x_i \vartriangleleft z_i)))).
\]

Theorem 8.1. Let \( (\Sigma,M) \) be an OP alphabet. \( L \) is accepted by a nondeterministic \( \omega \)OPBA \( \mathcal{A} \) over \( (\Sigma,M) \) if and only there exists an MSO sentence \( \varphi \) such that \( L = L^\omega(\varphi) \).

The construction of a nondeterministic \( \omega \)OPBA equivalent to an MSO formula is identical to the one given for finite strings.

The converse construction also follows essentially the same path as in the case of finite-length languages; thus, we only point out the few relevant differences w.r.t. the
construction of section 4. Formula $\varphi$ is defined as

$$
\varphi := \exists A_0, A_1, \ldots, A_N \bigg( \bigvee_{q_i \in I} \text{Start}_i \land \varphi_\delta \land \varphi_{\text{unique}} \land \bigvee_{q_f \in F} \text{Accept}_f \bigg),
$$

where $\text{Start}_i$ is defined as in section 4, and $\text{Accept}_f$ is a shortcut representing the Büchi acceptance condition (a final state is reached infinitely often):

$$
\text{Accept}_f := \forall x \exists y (x < y \land y \in Q_f).
$$

Formula $\varphi_\delta$ encodes the nondeterministic transition functions of the automaton and is obtained from formula $\varphi_{\text{push}} \land \varphi_{\text{halt}} \land \varphi_{\text{pop}}$ defined in section 4, by replacing expressions such as $q_k = \delta(\ldots)$ by expressions such as $q_k \in \delta(\ldots)$. Finally, formula $\varphi_{\text{unique}}$ is defined as the conjunction of the following formulas:

$$
\varphi_{\text{unique}A} := \forall x \bigg( \bigwedge_{i=0}^{N} (x \in A_i \Rightarrow \neg \bigvee_{j=0}^{N} (j \neq i \land x \in A_j)) \bigg),
$$

$$
\varphi_{\text{unique}\_\text{next}} := \forall x, y \bigg( \bigwedge_{k=0}^{N} (\text{Next}_k(x, y) \Rightarrow \neg \bigvee_{j \neq k} \text{Next}_j(x, y)) \bigg).
$$

Such formulas were not necessary in the finite case because they were implied by the determinism of the automaton.

The proof that formula $\varphi$ is satisfied by all and only the words accepted by $A$ is again based on Lemmas 4.3 and 4.4, but we need some more properties to cope with infinite words.

Any $\omega$-word $w \in \Sigma^\omega$ compatible with $M$ can be factored, as in the proof of Theorem 7.7, as a sequence $w = w_1 w_2 w_3 \ldots$, where either $w_i \in \Sigma$ is a pending letter or $w_i$ is the body of the chain $a_i[w_i]^b_i$, where $a_i$ is the last pending letter before $w_i$ and $b_i$ is the first symbol of $w_i+1$. A similar factorization holds for a finite word $\#w$ without end delimiter. We denote by $P$ the set of positions in a (finite or infinite) string $w$ that correspond to pending letters and by $E$ the set of positions of the right delimiter of the chains of the factorization. These two sets are not necessarily disjoint, and $EP$ is their union:

$$
z \in P := \forall x, y \ (x < z < y \land x \lor y \Rightarrow \#(y)),
$$

$$
z \in E := \exists x \ (x \in P \land x \lor z),
$$

$$
z \in EP := z \in P \lor z \in E.
$$

Any prefix of an infinite string $w$ which ends in an EP position of $w$ is called an $EP$-prefix of $w$.

Let us define

$$
\psi_{i,k}(A_0, \ldots, A_N, B_0, \ldots B_N, C_0, \ldots, C_N) := \text{Start}_i \land \varphi_\delta \land \text{Final}_k,
$$

where

$$
\text{Final}_k := \exists y \exists e \ (y \in Q_k \land y \leq e \land e \in EP \land \forall z (y \leq z \land z \in EP \Rightarrow z = e)).
$$
and $\varphi_i'$ is as $\varphi_a$ except for the formula $\varphi_{\delta_{pop}}$, where the constraint $\neg\#(y)$ is conjuncted to the antecedent of $\varphi_{\delta_{pop}w_w}$, and $\varphi_{\delta_{pop}w_wB}$ and $\varphi_{\delta_{pop}w_wC}$ are replaced by the unique formula

$$\varphi_{pop,w_w} := \forall x, z, v, y \left( \bigwedge_{k=0}^{N} \left( x \in B_k \land v \in C_k \land \neg\#(y) \land \text{Tree}(x, z, v, y) \right) \Rightarrow \bigvee_{i=0}^{N} \bigvee_{j=0}^{N} \left( \delta_{pop}(q_i, q_j) \ni q_k \right) \right).$$

We will interpret formula $\psi_{i,k}$ over finite strings. More precisely, let $w'$ be an EP-prefix of a string $w \in \Sigma^\omega$. It is $w \models_w \varphi$ iff there exist an initial state $q_i$, a final state $q_f$, and an assignment $A_{0}, \ldots, C_N$ such that $w' \models \psi_{i,f}(A_0, \ldots, C_N)$ for an infinite number of EP-prefixes $w'$ of $w$. In this case, a position $x$ in a prefix $w'$ may start a chain that goes beyond the end of $w'$, hence, in such cases $x$ is in $B_k$ in the assignment satisfying $w \models_w \varphi$ but $w' \not\models_{\omega} \varphi_{pop,w_wB}$. This is the reason why we replace the backward formulas of $\varphi_{\delta_{pop}}$ in $\varphi_i$.

For any assignment for $A_{0}, \ldots, C_N$, it is $w' \models \psi_{i,k}(A_0, \ldots, C_N)$ iff there exists a run of $A$ for $w'$ beginning from state $q_i$ that visits state $q_k$ somewhere after the last EP position before $|w'|$. The run can be built reasoning as in Lemmas 4.3 and 4.4 within the chains of the factorization, and using formulas $\varphi_{\delta_{push}}$ and $\varphi_{\delta_{shift}}$ for the positions of pending letters. The properties corresponding to states $q_i$ and $q_k$ are provided by formulas $\text{Start}_i$ and $\text{Final}_k$. If $w'$ and $w''$ are EP-prefixes of $w$ and both satisfy $\psi_{i,k}$ with the same assignment to $A_{0}, \ldots, C_N$, then for the corresponding runs built with such a construction, one is the prefix of the other.

Hence $w \models_w \varphi$ if and only if there exist infinitely many (finite) runs of $A$ on EP-prefixes of $w$, each of them beginning from $q_i$ and visiting the same final state $q_f$ somewhere after its last EP position; such runs are all prefixes of the same infinite run $\rho$.

Furthermore, since there is a move in $\rho$ that reaches $q_f$ while reading the suffix of each of those EP-prefixes after its last EP position, then $\rho$ traverses infinitely often $q_f$, and hence $\rho$ is accepting for $A$.

Symmetrically, one can prove that if there exists an accepting run $\rho$ for an $\omega$-string $w$ in $A$, then $w \models_w \varphi$.

9. **Concluding remarks.** In this paper we have supplied a number of results about OPLs which, together with previous recent and less recent ones, qualify OPLs as the largest class of deterministic CF languages that enjoy all of the following basic properties which have a strong impact on various types of practical applications, spanning from parsing to model checking:

- Local parsability: this property, not pursued in this paper, allows for realizing simple and efficient parallel and/or incremental algorithms [7, 6].
- Closure under all main language operations—Boolean ones, concatenation, Kleene *, and others [18].
- Automata-theoretic and MSO logic characterization.
- Extension of all the above properties to the case of $\omega$-languages, i.e., languages consisting of infinite-length strings, with the noticeable and typical exception of the lack of equivalence between deterministic and nondeterministic automata—under the Büchi acceptance condition.

As for the complexity of the constructions used to prove our results we have shown that they are in general of the same order as those of corresponding constructions for less powerful language families—typically, VPLs; the few cases of different complexity have been pointed out in Table 8.
This fairly complete foundational characterization of OPLs can now ignite—and partially already did—further research along several directions. On the one side we are developing practical tools exploiting the above properties both in parsing and in automatic verification; on the other side we envisage many interesting special cases of OPLs motivated by different possible applications.

For instance, we are investigating the use of logic formalisms simpler than MSO logic to characterize suitable subclasses of general OPLs, in the same vein as it has been done for regular languages [15], VPLs [2], and for various cases of tree languages [1, 10]; a first result on this respect is that free languages, a subclass of OPLs originally motivated by grammar inference [19, 20] can be defined in terms of a first-order logic rather than a second-order one [27].

We are also investigating new, less usual application fields for OPLs, or suitable subclasses thereof, beyond the traditional field of programming languages, e.g., in the direction suggested by Examples 5, 8, and others not reported here which are in the same vein as the application indicated for VPLs but considerably extend its scope.

Acknowledgment. We are very grateful to the anonymous reviewers for their constructive criticism and their thoughtful, precise, and very detailed suggestions to improve the original manuscript.

REFERENCES