

# Beyond $\omega BS$ -regular Languages: $\omega T$ -regular Expressions and Counter-Check Automata\*

Dario Della Monica

Universidad Complutense de Madrid, Spain, and  
Università “Federico II” di Napoli, Italy.  
ddellamo@ucm.es

Angelo Montanari

Università di Udine, Italy.  
angelo.montanari@uniud.it

Pietro Sala

Università di Verona, Italy.  
pietro.sala@univr.it

In the last years, various extensions of  $\omega$ -regular languages have been proposed in the literature, including  $\omega B$ -regular ( $\omega$ -regular languages extended with boundedness),  $\omega S$ -regular ( $\omega$ -regular languages extended with strict unboundedness), and  $\omega BS$ -regular languages (the combination of  $\omega B$ - and  $\omega S$ -regular ones). While the first two classes satisfy a generalized closure property, namely, the complement of an  $\omega B$ -regular (resp.,  $\omega S$ -regular) language is an  $\omega S$ -regular (resp.,  $\omega B$ -regular) one, the last class is not closed under complementation. The existence of non- $\omega BS$ -regular languages that are the complements of some  $\omega BS$ -regular ones and express fairly natural properties of reactive systems motivates the search for other well-behaved classes of extended  $\omega$ -regular languages. In this paper, we introduce the class of  $\omega T$ -regular languages, that includes meaningful languages which are not  $\omega BS$ -regular. We first define it in terms of  $\omega T$ -regular expressions. Then, we introduce a new class of automata (counter-check automata) and we prove that (i) their emptiness problem is decidable in PTIME and (ii) they are expressive enough to capture  $\omega T$ -regular languages (whether or not  $\omega T$ -regular languages are expressively complete with respect to counter-check automata is still an open problem). Finally, we provide an encoding of  $\omega T$ -regular expressions into S1S+U.

## 1 Introduction

A fundamental role in computer science is played by  $\omega$ -regular languages, as they provide a natural setting for the specification and verification of nonterminating finite-state systems. Since the seminal work by Büchi [7], McNaughton [13], and Elgot and Rabin [9] in the sixties, a great research effort has been devoted to the theory and the applications of  $\omega$ -regular languages. Equivalent characterisations of  $\omega$ -regular languages have been given in terms of formal languages ( $\omega$ -regular expressions), automata (Büchi, Rabin, and Muller automata), classical logic (weak/strong monadic second-order logic of one successor,  $\omega S1S/S1S$  for short), and temporal logic (Quantified Linear Temporal Logic, Extended Temporal Logic).

Recently, it has been shown that  $\omega$ -regular languages can be extended in various ways, preserving their decidability and some of their closure properties [4, 5]. As an example, extended  $\omega$ -regular languages make it possible to constrain the distance between consecutive occurrences of a given symbol to be (un)bounded (in the limit). Boundedness comes into play in the study of *finitary fairness* as opposed to the classic notion of *fairness*, widely used in automated verification of concurrent systems. According to the latter, no individual process in a multi-process system may be ignored for ever; finitary fairness imposes the stronger constraint that every enabled transition is executed within at most  $b$  time-units, where  $b$  is an unknown, constant bound. In [1], it is shown that finitary fairness enjoys some desirable mathematical properties that are violated by the weaker notion of fairness, and yet it captures all reasonable

---

\*This work was partially supported by the Italian INdAM-GNCS project *Logics and Automata for Interval Model Checking*. In addition, D. Della Monica acknowledges the financial support from a Marie Curie INdAM-COFUND-2012 Outgoing Fellowship.

schedulers' implementations. The same property has been investigated from a logical perspective in [11], where the logic PROMPT-LTL is introduced. Roughly speaking, PROMPT-LTL extends LTL with the *prompt-eventually* operator, which states that an event will happen within the next  $b$  time-units,  $b$  being an unknown, constant bound. An analogous extension has been recently proposed for the propositional interval logic of temporal neighborhood PNL [8].

From the point of view of formal languages, the proposed extensions pair the Kleene star  $(\cdot)^*$  with bounding/unbounding variants of it. Intuitively, the bounding exponent  $(\cdot)^B$  (aka  $B$ -constructor) constrains parts of the input word to be of bounded size, while the unbounding exponent  $(\cdot)^S$  (aka  $S$ -constructor) forces parts of the input word to be arbitrarily large. The two extensions have been studied both in isolation ( $\omega B$ - and  $\omega S$ -regular expressions) and in conjunction ( $\omega BS$ -regular expressions). Equivalent characterisations of extended  $\omega$ -regular languages are given in [4, 5] in terms of automata ( $\omega B$ -,  $\omega S$ -, and  $\omega BS$ -automata) and classical logic (fragments of  $\omega S1S+U$ , i.e., the extension of  $\omega S1S$  with the unbounding quantifier  $\mathbb{U}$  [3], that allows one to express properties which are satisfied by finite sets of arbitrarily large size).<sup>1</sup> In [5], the authors also show that the complement of an  $\omega B$ -regular language is an  $\omega S$ -regular one and vice versa; moreover, they show that  $\omega BS$ -regular languages, featuring both  $B$ - and  $S$ -constructors, strictly extend  $\omega B$ - and  $\omega S$ -regular languages and are not closed under complementation.

In this paper, we focus on those  $\omega$ -languages which are complements of  $\omega BS$ -regular ones, but are not  $\omega BS$ -regular. We start with an in-depth analysis of a paradigmatic example of one such language [5]. It allows us to identify a meaningful extension of  $\omega$ -regular languages ( $\omega T$ -regular languages) including it and obtained by adding a new, fairly natural constructor  $(\cdot)^T$ , named  $T$ -constructor, to the standard constructors of  $\omega$ -regular expressions. An interesting feature of such a class is that pairing  $(\cdot)^B$  and  $(\cdot)^S$  with  $(\cdot)^T$  one can capture all possible ways of instantiating  $*$ -expressions (this is not the case with  $B$  and  $S$  only). In view of that, it can be said that  $(\cdot)^T$  “complements”  $(\cdot)^B$  and  $(\cdot)^S$  with respect to  $(\cdot)^*$ . Then, we introduce a new class of automata (counter-check automata), that are expressive enough to capture  $\omega T$ -regular languages, and we show that their emptiness problem is decidable. Finally, we provide an encoding of  $\omega T$ -regular expressions (languages) into  $S1S+U$ .

The paper is organized as follows. In Section 2, we illustrate existing extensions of  $\omega$ -regular languages, with a special attention to  $\omega BS$ -regular ones, and we introduce the class of  $\omega T$ -regular languages. In Section 3, we define counter-check automata (CCA) and prove that their emptiness problem is decidable in PTIME. In Section 4, we provide an encoding of  $\omega T$ -regular languages into CCA, while, in Section 5, we show that they can be defined in  $S1S+U$ . Conclusions provide an assessment of the work done and outline future research directions.

## 2 Extensions of $\omega$ -regular languages

In this section, we give a short account of the extensions of  $\omega$ -regular languages proposed in the literature (details can be found in [4, 5]) and we outline a new one. To begin with, we observe that an  $\omega$ -word can be seen as the concatenation of a finite prefix, belonging to a regular language, and an infinite sequence of finite words (we call each of these finite words an  $\omega$ -iteration), also belonging to a regular language. A standard way to define  $\omega$ -regular languages is by means of  $\omega$ -regular expressions. An interesting case is that of  $\omega$ -iterations consisting of a finite sequence of words, generated by an occurrence of the Kleene star operator  $(\cdot)^*$ , aka  $*$ -constructor, in the scope of the  $\omega$ -constructor  $(\cdot)^\omega$ . As an example, the  $\omega$ -regular expression  $(a^*b)^\omega$  generates the language of  $\omega$ -words featuring an infinite sequence of  $\omega$ -iterations, each one consisting of a finite (possibly empty) sequence of  $a$ 's followed by exactly one  $b$ .

---

<sup>1</sup>Undecidability of full  $S1S+U$  has been shown in [6].

Given an  $\omega$ -regular expression  $E$  featuring an occurrence of  $(\cdot)^*$  (sub-expression  $R^*$ ) in the scope of  $(\cdot)^\omega$  and an  $\omega$ -word  $w$  belonging to the language of  $E$ , we refer to the sequence of the sizes of the (maximal) blocks of consecutive iterations of  $R$  in the different  $\omega$ -iterations as the *(sequence of) exponents of  $R$  in (the  $\omega$ -iterations of)  $w$* . Let  $w = abaabaaaab\dots$  be an  $\omega$ -word generated by the  $\omega$ -regular expression  $(a^*b)^\omega$ . The sequence of exponents of  $a$  in  $w$  is  $1, 2, 3, \dots$ . Sometimes, we will denote words in a compact way, by explicitly indicating the exponents of a sub-expression, e.g., we will write  $w$  as  $a^1ba^2ba^3b\dots$ .

Given an expression  $E$ , we denote by  $\mathcal{L}(E)$  the language defined by  $E$ . With a little abuse of notation, we will sometimes identify a language with the expression defining it, and vice versa, e.g., we will write “language  $(a^*b)^\omega$ ” instead of “language  $\mathcal{L}((a^*b)^\omega)$ ”. Notice that  $(\cdot)^*$  allows one to impose the existence of a finite sequence of words (described by its argument expression) within each  $\omega$ -iteration, but it cannot be used to express properties of the sequence of exponents of its argument expression in the  $\omega$ -iterations of an  $\omega$ -word. To overcome such a limitation, some meaningful extensions of  $\omega$ -regular expressions have been investigated in the last years, that make it possible to constrain the behavior of  $(\cdot)^*$  in the limit.

**Beyond  $\omega$ -regularity.** A first class of extended  $\omega$ -regular languages is that of  $\omega B$ -regular languages, that allow one to impose boundedness conditions.  $\omega B$ -regular expressions are obtained from  $\omega$ -regular ones by adding a variant of  $(\cdot)^*$ , called *B-constructor* and denoted by  $(\cdot)^B$ , to be used in the scope of  $(\cdot)^\omega$ . The bounded exponent  $B$  allows one to constrain the argument  $R$  of the expression  $R^B$  to be repeated in each  $\omega$ -iteration a number of times less than a certain bound fixed for the whole  $\omega$ -word. As an example, the expression  $(a^Bb)^\omega$  denotes the language of  $\omega$ -words in  $(a^*b)^\omega$  for which there is an upper bound on the number of consecutive occurrences of  $a$  (the sequence of exponents of  $a$  is bounded). As the bound may vary from word to word, the language is not  $\omega$ -regular. The class of  $\omega S$ -regular languages extends that of  $\omega$ -regular ones with strong unboundedness. By analogy with  $\omega B$ -regular expressions,  $\omega S$ -regular expressions are obtained from  $\omega$ -regular ones by adding a variant of  $(\cdot)^*$ , called *S-constructor* and denoted by  $(\cdot)^S$ , to be used in the scope of  $(\cdot)^\omega$ . For every  $\omega S$ -regular expression containing the sub-expression  $R^S$  and every natural number  $k > 0$ , the strictly unbounded exponent  $S$  constrains the number of  $\omega$ -iterations in which the argument  $R$  is repeated at most  $k$  times to be finite. Let us consider  $\omega$ -words that feature an infinite number of instantiations of the expression  $R^S$ , that is,  $\omega$ -words for which there exists an infinite number of  $\omega$ -iterations including a sequence of consecutive  $R$ 's generated by  $R^S$ . It can be easily checked that in these words the sequence of exponents of  $R$  tends towards infinity. As an example, the expression  $(a^Sb)^\omega$  denotes the language of  $\omega$ -words  $w$  in  $(a^*b)^\omega$  such that, for any  $k > 0$ , there exists a suffix of  $w$  that only features maximal sequences of consecutive  $a$ 's that are longer than  $k$ .

$\omega BS$ -regular expressions are built by using the operators of  $\omega$ -regular expressions and both  $(\cdot)^B$  and  $(\cdot)^S$ . In [5], the authors show that the class of  $\omega BS$ -regular languages strictly includes the classes of  $\omega B$ - and  $\omega S$ -regular languages as witnessed by the  $\omega BS$ -regular language  $L = (a^Bb + a^Sb)^\omega$  consisting of those  $\omega$ -words  $w$  featuring infinitely many occurrences of  $b$  and such that there are only finitely many numbers occurring infinitely often in the sequence of exponents of  $a$  in  $w$ , that is, there is a bound  $k$  such that no  $h > k$  occurs infinitely often in the sequence of exponents of  $a$  in  $w$ .  $L$  is neither  $\omega B$ - nor  $\omega S$ -regular.<sup>2</sup> Moreover, they prove that the class of  $\omega BS$ -regular languages is not closed under complementation. A counterexample is given precisely by  $L$ , whose complement is not  $\omega BS$ -regular (notice that  $\omega BS$ -regular languages whose complement is not an  $\omega BS$ -regular language are neither  $\omega B$ - nor  $\omega S$ -regular languages, as the complement of an  $\omega B$ -regular language is an  $\omega S$ -regular one and vice versa).

In this paper, we investigate those  $\omega$ -languages that do not belong to the class of  $\omega BS$ -regular languages, but whose complement belongs to this class. Let us consider, for instance, the complement  $\bar{L}$  of

<sup>2</sup>The constructor  $+$  occurring in  $L$  must not be thought of as performing the union of two languages, but rather as a “shuffling operator” that mixes  $\omega$ -iterations belonging to the two different (sub-)languages.

the language  $L$  above. Any word  $w$  in  $\bar{L}$  that features infinitely many occurrences of  $b$  (i.e.,  $w \in (a^*b)^\omega$ ) is such that there are infinitely many natural numbers that occur infinitely often in the sequence of exponents of  $a$  in  $w$ . By way of contradiction, suppose that there are only finitely many. Let  $k$  be the largest one. Now,  $w$  can be viewed as an infinite sequence of  $\omega$ -iterations, each of them characterised by the corresponding exponent of  $a$ . If the exponent associated with an  $\omega$ -iteration is greater than  $k$ , then it does not occur infinitely often, and thus the  $\omega$ -iteration is captured by the sub-expression  $a^S b$ . Otherwise, if the exponent is not greater than  $k$ , then the corresponding  $\omega$ -iteration is captured by the sub-expression  $a^B b$ . As an example, the  $\omega$ -word  $a^1 b a^2 b a^1 b a^3 b a^1 b a^4 b \dots$  does not belong to  $\bar{L}$  as 1 is the only exponent occurring infinitely often, while the  $\omega$ -word  $a^1 b a^2 b a^1 b a^2 b a^3 b a^1 b a^2 b a^3 b a^4 b \dots$  does belong to it as infinitely many (actually all) natural numbers occur infinitely often in the sequence of exponents.

Here, we focus on  $\omega$ -words featuring infinitely many exponents occurring infinitely often. More precisely, we introduce a new variant of  $(.)^*$ , called *T-constructor* and denoted by  $(.)^T$ , to be used in the scope of  $(.)^\omega$ , and we define the corresponding class of extended  $\omega$ -regular languages ( $\omega$ T-regular languages). Let  $E$  be an  $\omega$ -expression and let  $w \in E$ . An expression  $R^T$  occurring in  $E$  forces the sequence of exponents in  $w$  to feature infinitely many different elements occurring infinitely often. As an example, it can be easily checked that the language  $\bar{L}$  can be defined as  $(a^T b)^\omega + (a^* b^*)^* a^\omega$ , and thus it belongs to the class of  $\omega$ T-regular languages. In the following, we first provide a formal account of  $\omega$ BS-regular languages [5] and then we define  $\omega$ T-regular ones.

**$\omega$ BS-regular languages.** The class of  *$\omega$ BS-regular languages* is the class of languages defined by  $\omega$ BS-regular expressions. These latter are built on top of BS-regular expressions, just as  $\omega$ -regular expressions are built on top of regular ones. Let  $\Sigma$  be a finite, non-empty alphabet. A *BS-regular expression* over  $\Sigma$  is defined by the grammar [5]:

$$e ::= \emptyset \mid a \mid e \cdot e \mid e + e \mid e^* \mid e^B \mid e^S$$

with  $a \in \Sigma$ . Sometimes we omit the concatenation operator, thus writing  $ee$  for  $e \cdot e$ .

BS-regular expressions differ from standard regular ones for the presence of the constructors  $(.)^B$  and  $(.)^S$ . Since these operators constrain the behavior of the sequence of  $\omega$ -iterations to the limit, it is not possible to simply define the semantics of BS-regular expressions in terms of languages of (finite) words, and then to obtain  $\omega$ BS-regular languages through infinitely many, unrelated iterations of such words. Instead, we specify their semantics in terms of languages of infinite sequences of finite words; suitable constraints are imposed to such sequences in order to capture the intended meaning of  $(.)^B$  and  $(.)^S$ .

Let  $\mathbb{N}$  be the set of natural numbers, including 0, and  $\mathbb{N}_{>0} = \mathbb{N} \setminus \{0\}$ . For an infinite sequence  $\vec{u}$  of finite words over  $\Sigma$ , we denote by  $u_i$  ( $i \in \mathbb{N}_{>0}$ ) its  $i$ -th element. The semantics of BS-regular expressions over  $\Sigma$  is defined as follows (hereafter we assume  $f(0) = 1$ ):

- $\mathcal{L}(\emptyset) = \emptyset$ ;
- for  $a \in \Sigma$ ,  $\mathcal{L}(a)$  only contains the infinite sequence of the one-letter word  $a \{(a, a, a, \dots)\}$ ;
- $\mathcal{L}(e_1 \cdot e_2) = \{\vec{w} \mid \forall i. w_i = u_i \cdot v_i, \vec{u} \in \mathcal{L}(e_1), \vec{v} \in \mathcal{L}(e_2)\}$ ;
- $\mathcal{L}(e_1 + e_2) = \{\vec{w} \mid \forall i. w_i \in \{u_i, v_i\}, \vec{u}, \vec{v} \in \mathcal{L}(e_1) \cup \mathcal{L}(e_2)\}$ ;<sup>3</sup>
- $\mathcal{L}(e^*) = \{(u_{f(0)} u_2 \dots u_{f(1)-1}, u_{f(1)} \dots u_{f(2)-1}, \dots) \mid \vec{u} \in \mathcal{L}(e) \text{ and } f : \mathbb{N} \rightarrow \mathbb{N}_{>0} \text{ is an unbounded and nondecreasing function}\}$ ;
- $\mathcal{L}(e^B) = \{(u_{f(0)} u_2 \dots u_{f(1)-1}, u_{f(1)} \dots u_{f(2)-1}, \dots) \mid \vec{u} \in \mathcal{L}(e) \text{ and } f : \mathbb{N} \rightarrow \mathbb{N}_{>0} \text{ is an unbounded and nondecreasing function such that } \exists n \in \mathbb{N} \forall i \in \mathbb{N}. (f(i+1) - f(i) < n)\}$ ;
- $\mathcal{L}(e^S) = \{(u_{f(0)} u_2 \dots u_{f(1)-1}, u_{f(1)} \dots u_{f(2)-1}, \dots) \mid \vec{u} \in \mathcal{L}(e) \text{ and } f : \mathbb{N} \rightarrow \mathbb{N}_{>0} \text{ is an unbounded}$

<sup>3</sup>Unlike the case of word languages, when applied to languages of word sequences, the operator  $+$  does not return the union of the two argument languages. As an example,  $\mathcal{L}(a) \cup \mathcal{L}(b) \subsetneq \mathcal{L}(a+b)$ , as witnessed by the word sequence  $(a, b, a, b, a, b, \dots)$ . In general, for all BS-regular expressions  $e_1, e_2$ , it holds that  $\mathcal{L}(e_1) \cup \mathcal{L}(e_2) \subseteq \mathcal{L}(e_1 + e_2)$ .

and nondecreasing function such that  $\forall n \in \mathbb{N} \exists k \in \mathbb{N} \forall i > k. (f(i+1) - f(i) > n)$ .

Given a sequence  $\vec{v} = (u_{f(0)}u_2 \dots u_{f(1)-1}, u_{f(1)} \dots u_{f(2)-1}, \dots) \in e^{op}$ , where  $op \in \{*, B, S\}$ , we formally define the *sequence of exponents of  $e$  in  $\vec{v}$* , denoted by  $N(\vec{v})$ , as the sequence  $(f(i+1) - f(i))_{i \in \mathbb{N}}$ . While the  $*$ -constructor does not impose any constraint on the sequence of exponents of its operand, the  $B$ -constructor forces the sequence of exponents to be bounded and the  $S$ -constructor forces it to be strictly unbounded, that is, its limit inferior tends towards infinity (equivalently, the  $S$ -constructor imposes that no exponent occurs infinitely many times in the sequence).

The  $\omega$ -constructor defines languages of infinite words from languages of infinite word sequences. Let  $e$  be a  $BS$ -regular expression. The semantics of the  $\omega$ -constructor is defined as follows:

- $\mathcal{L}(e^\omega) = \{w \mid w = u_1u_2u_3 \dots \text{ for some } \vec{u} \in \mathcal{L}(e)\}$ .

$\omega BS$ -expressions are defined by the grammar (we denote languages of word sequences by lowercase letters, such as  $e, e_1, \dots$ , and languages of words by uppercase ones, such as  $E, E_1, \dots, R, R_1, \dots$ ):

$$E ::= E + E \mid R \cdot E \mid e^\omega$$

where  $R$  is a regular expression,  $e$  is a  $BS$ -regular expression, and  $+$  and  $\cdot$  respectively denote union and concatenation of word languages (formally,  $\mathcal{L}(E_1 + E_2) = \mathcal{L}(E_1) \cup \mathcal{L}(E_2)$  and  $\mathcal{L}(E_1 \cdot E_2) = \{u \cdot v \mid u \in \mathcal{L}(E_1), v \in \mathcal{L}(E_2)\}$ ).<sup>4</sup> As we did in the case of languages of word sequences, we will sometimes omit the concatenation operator between word languages.

**$\omega T$ -regular languages.** We are now ready to introduce  $\omega T$ -regular languages. From [5], we know that the class of  $\omega BS$ -regular languages is not closed under complementation, that is, there are  $\omega$ -languages that are the complements of  $\omega BS$ -regular ones while being not  $\omega BS$ -regular. This is the case, for instance, with the complement  $\bar{L}$  of the  $\omega BS$ -regular language  $L = (a^B b + a^S b)^\omega$ . We have already pointed out the distinctive features of  $\bar{L}$ , showing that  $\omega$ -words belonging to it are, to a certain extent, characterised by sequences of exponents where infinitely many exponents occur infinitely often. In order to capture extended  $\omega$ -regular languages that satisfy such a property, we define a new class of  $\omega$ -regular languages, called  *$\omega T$ -regular languages*. It includes those languages that can be expressed by  *$\omega T$ -regular expressions*, which are defined by the grammar (where  $R$  is a regular expression and  $a \in \Sigma$ ):

$$\begin{aligned} E &::= E + E \mid R \cdot E \mid e^\omega \\ e &::= \emptyset \mid a \mid e \cdot e \mid e + e \mid e^* \mid e^T \end{aligned}$$

The sub-grammar rooted in the non-terminal  $e$  generates the  *$T$ -regular expressions*. The only new ingredient in the above definition is the  $T$ -constructor  $(\cdot)^T$ , that, given a language of word sequences  $e$ , defines the following language:

- $\mathcal{L}(e^T) = \{(u_{f(0)}u_2 \dots u_{f(1)-1}, u_{f(1)} \dots u_{f(2)-1}, \dots) \mid \vec{u} \in \mathcal{L}(e) \text{ and } f : \mathbb{N} \rightarrow \mathbb{N}_{>0} \text{ is an unbounded and nondecreasing function such that } \exists^\omega n \in \mathbb{N} \forall k \in \mathbb{N} \exists i > k. (f(i+1) - f(i) = n)\}$ ,

where  $\exists^\omega$  is a shorthand for “there are infinitely many”.

For  $\vec{u} \in e^T$ , we define the *sequence of exponents of  $e$  in  $\vec{u}$* , denoted by  $N(\vec{u})$ , exactly as we did in the case of  $BS$ -regular expressions. Moreover, for  $op \in \{*, B, S, T\}$  and  $\vec{u} \in e^{op}$ , we denote by  $N_i(\vec{u})$  (resp.,  $N_f(\vec{u})$ ) the set of exponents occurring infinitely (resp., finitely) many times in  $N(\vec{u})$ . It is not difficult to see that the cardinality of  $N_i(\vec{u})$  is infinite, for every  $\vec{u} \in e^T$ , and thus the formal semantics of the  $T$ -constructor conforms with the intuitive one given at the end of Subsection 2.

It is not difficult to devise an  $\omega T$ -regular language that is not  $\omega BS$ -regular and, vice versa, of an  $\omega BS$ -regular language that is not  $\omega T$ -regular.

As we already pointed out in the introduction, one of the motivations for the proposal of the  $T$ -constructor stems from the fact that it somehow complements the other two with respect to the Kleene star.

<sup>4</sup>Notice the abuse of notation with the previous definition of the operators  $+$  and  $\cdot$  over languages of word sequences.

We can make such a claim more precise as follows. Let  $\vec{u} \in \mathcal{L}(e^{op})$ , with  $op \in \{B, S, T\}$ . If  $\vec{u} \in \mathcal{L}(e^B)$ , then  $N(\vec{u})$  is bounded, while if either  $\vec{u} \in \mathcal{L}(e^S)$  or  $\vec{u} \in \mathcal{L}(e^T)$  it is unbounded; moreover, if  $\vec{u} \in \mathcal{L}(e^S)$ , then  $N_i(\vec{u}) = \emptyset$ , while if  $\vec{u} \in \mathcal{L}(e^T)$ , then  $N_i(\vec{u})$  is infinite. The next proposition shows that when paired with  $(\cdot)^B$  and  $(\cdot)^S, (\cdot)^T$  makes it possible to define the Kleene star. Let *BST-regular expressions* be obtained from *BS-regular* ones by enriching them with  $(\cdot)^T$ .

**Proposition 1.** *For every BST-regular expression  $e$ , it holds that  $e^* = e^B + e^S + e^T$ .*

*Proof.* As  $\mathcal{L}(e^B) \subseteq \mathcal{L}(e^*)$ ,  $\mathcal{L}(e^S) \subseteq \mathcal{L}(e^*)$ , and  $\mathcal{L}(e^T) \subseteq \mathcal{L}(e^*)$ , it trivially holds that  $\mathcal{L}(e^B + e^S + e^T) \subseteq \mathcal{L}(e^*)$ .

To prove the converse inclusion, we assume that  $\vec{v} \in \mathcal{L}(e^*)$  and we show that  $\vec{v} \in \mathcal{L}(e^B + e^S + e^T)$ . By the semantics of  $e^*$ ,  $\vec{v} = (u_{f(0)}u_{f(1)} \dots u_{f(1)-1}, u_{f(1)} \dots u_{f(2)-1}, \dots)$ , for a word sequence  $\vec{u} \in \mathcal{L}(e)$  and an unbounded and nondecreasing function  $f: \mathbb{N} \rightarrow \mathbb{N}_{>0}$ , with  $f(0) = 1$ .

Let  $N(\vec{v})$  be the sequence of exponents  $(n_1, n_2, \dots)$ . If  $N(\vec{v})$  is bounded, then  $\vec{v} \in \mathcal{L}(e^B) \subseteq \mathcal{L}(e^B + e^S + e^T)$ . Otherwise, let  $I = \langle i_1, i_2, \dots \rangle$  be the increasing sequence of indexes  $i_j$  such that  $n_{i_j} \in N_i(\vec{v})$  and  $F = \langle f_1, f_2, \dots \rangle$  be the increasing sequence of indexes  $f_j$  such that  $n_{f_j} \in N_{f_j}(\vec{v})$ . It clearly holds that  $I \cup F = \mathbb{N}_{>0}$ . Now, let  $\vec{t} = \langle t_1, t_2, \dots \rangle$  be the word sequence such that  $t_{f_j} = v_{f_j}$  for every  $f_j \in F$  and  $t_{i_j} = u_1 \dots u_{i_j}$  for every  $i_j \in I$ . Clearly,  $\vec{t} \in \mathcal{L}(e^S)$ . Moreover, let  $\vec{w} = \langle w_1, w_2, \dots \rangle$  be the word sequence such that  $w_{i_j} = v_{i_j}$  for every  $i_j \in I$  and  $w_{f_j} = v_{i_1}$  for every  $f_j \in F$ . If  $N_i(\vec{v})$  is finite, then  $N(\vec{w})$  is bounded by  $\max(N_i(\vec{v}))$ , and thus  $\vec{w} \in \mathcal{L}(e^B)$ ; otherwise,  $N_i(\vec{w}) (= N_i(\vec{v}))$  is infinite, that is, there are infinitely many exponents in  $\vec{w}$  occurring infinitely often, and thus  $\vec{w} \in \mathcal{L}(e^T)$ . Hence,  $\vec{w} \in \mathcal{L}(e^B) \cup \mathcal{L}(e^T) \subseteq \mathcal{L}(e^B + e^T)$ . Since  $\vec{v}$  is such that  $v_k = t_k$ , if  $k \in F$ , and  $v_k = w_k$ , if  $k \in I$ ,  $\vec{v} \in \mathcal{L}(e^B + e^S + e^T)$ .  $\square$

### 3 Counter-check automata

In this section, we introduce a new class of automata, called counter-check automata, and we show that their emptiness problem is decidable in PTIME. In the next section, we will show that they are expressive enough to encode  $\omega$ T-regular expressions.

A *counter-check automaton* (an example is given in Figure 1) is an automaton equipped with a fixed number of counters. A transition can possibly *increment* or *reset* one of them (or do nothing). We refer to reset operations as *check* operations to put the emphasis on the fact that computations keep trace of the evolution of the counter values. In particular, the acceptance condition depends on the sequences of *check values* (i.e., the values when a check operation is performed) for all counters.

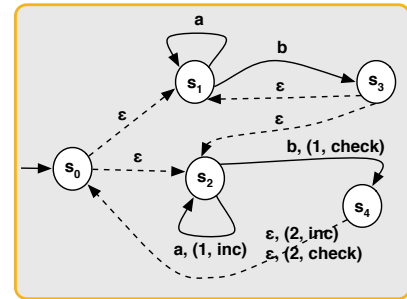


Figure 1: A CCA for the language  $((a^*b)^*a^T b)^\omega$  ( $N = 2$ ).

**Definition 1 (CCA).** A counter-check automaton (CCA for short) is a quintuple  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$ , where  $S$  is a finite set of states,  $\Sigma$  is a finite alphabet,  $s_0 \in S$  is the initial state,  $N \in \mathbb{N}_{>0}$  is the number of counters, and  $\Delta \subseteq S \times (\Sigma \cup \{\varepsilon\}) \times S \times (\{1, \dots, N\} \times \{no\_op, inc, check\})$  is a transition relation, subject to the constraint: if  $(s, \sigma, s', (k, op)) \in \Delta$  and  $op = no\_op$ , then  $k = 1$ .

A *configuration* of a CCA  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  is a pair  $(s, \mathbf{v})$ , where  $s \in S$  and  $\mathbf{v} \in \mathbb{N}^N$  is called *counter vector*. For  $\mathbf{v} \in \mathbb{N}^N$  and  $i \in \{1, \dots, N\}$ , let  $\mathbf{v}[i]$  be the  $i$ -th component of  $\mathbf{v}$ , i.e., the value of the  $i$ -th counter.

Let  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  be a CCA. We define a ternary relation  $\rightarrow_{\mathcal{A}}$  over pairs of configurations and symbols in  $\Sigma \cup \{\varepsilon\}$  such that for all configuration pairs  $(s, \mathbf{v}), (s', \mathbf{v}')$  and  $\sigma \in \Sigma \cup \{\varepsilon\}$ ,  $(s, \mathbf{v}) \rightarrow_{\mathcal{A}}^\sigma (s', \mathbf{v}')$  iff there is  $\delta = (s, \sigma, s', (k, op)) \in \Delta$  such that  $\mathbf{v}'[h] = \mathbf{v}[h]$  for all  $h \neq k$ , and

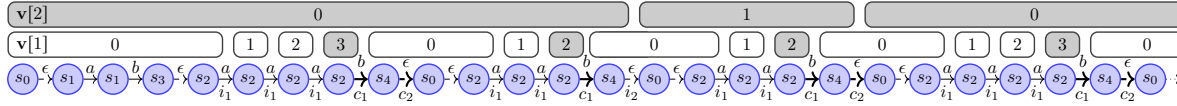


Figure 2: A prefix of a computation of the automaton in Figure 1. A configuration is characterised by a circle (state) and the rounded-corner rectangles above it (counter vector).  $\mathbf{v}[i]$  is a counter vector component. Checked values for counters are highlighted in gray, with the corresponding transitions being written in boldface.

- if  $op = no\_op$ , then  $\mathbf{v}'[k] = \mathbf{v}[k]$ ;
- if  $op = inc$ , then  $\mathbf{v}'[k] = \mathbf{v}[k] + 1$ ;
- if  $op = check$ , then  $\mathbf{v}'[k] = 0$ .

In such a case, we say that  $(s, \mathbf{v}) \rightarrow_{\mathcal{A}}^{\sigma} (s', \mathbf{v}')$  via  $\delta$ . Let  $\rightarrow_{\mathcal{A}}^*$  be the reflexive and transitive closure of  $\rightarrow_{\mathcal{A}}^{\sigma}$  (where we abstract away symbols in  $\Sigma \cup \{\varepsilon\}$ ). The *initial configuration* of  $\mathcal{A}$  is the pair  $(s_0, \mathbf{v}_0)$ , where for each  $k \in \{1, \dots, N\}$  we have  $\mathbf{v}_0[k] = 0$ . A *computation* of  $\mathcal{A}$  is an infinite sequence of configurations  $\mathcal{C} = (s_0, \mathbf{v}_0)(s_1, \mathbf{v}_1) \dots$ , where, for all  $i \in \mathbb{N}$ ,  $(s_i, \mathbf{v}_i) \rightarrow_{\mathcal{A}}^{\sigma_i} (s_{i+1}, \mathbf{v}_{i+1})$  for some  $\sigma_i \in \Sigma \cup \{\varepsilon\}$  (see Figure 2). For a computation  $\mathcal{C} = (s_0, \mathbf{v}_0)(s_1, \mathbf{v}_1) \dots$  we let  $check_{\mathcal{C}, k}^{\infty}$  ( $k \in \{1, \dots, N\}$ ) denote the set  $\{n \in \mathbb{N} \mid \forall h \exists i > h \text{ such that } \mathbf{v}_i[k] = n \text{ and } \mathbf{v}_{i+1}[k] = 0\}$ , that is,  $check_{\mathcal{C}, k}^{\infty}$  is the set of values of the  $k$ -th counter that are checked infinitely often along  $\mathcal{C}$ . Given two configurations  $(s_i, \mathbf{v}_i)$  and  $(s_j, \mathbf{v}_j)$  in  $\mathcal{C}$ , with  $i \leq j$ , we say that  $(s_j, \mathbf{v}_j)$  is  $\varepsilon$ -*reachable* from  $(s_i, \mathbf{v}_i)$ , written  $(s_i, \mathbf{v}_i) \rightarrow_{\mathcal{A}}^{*\varepsilon} (s_j, \mathbf{v}_j)$ , if  $(s_{j'-1}, \mathbf{v}_{j'-1}) \rightarrow_{\mathcal{A}}^{\varepsilon} (s_{j'}, \mathbf{v}_{j'})$  for all  $j' \in \{i+1, \dots, j\}$ .

A *run*  $\pi$  of  $w$  on  $\mathcal{A}$  is a computation  $\pi = (s_0, \mathbf{v}_0)(s_1, \mathbf{v}_1) \dots$  for which there exists an increasing function  $f : \mathbb{N}_{>0} \rightarrow \mathbb{N}$ , called *trace of  $w$  in  $\pi$  wrt.  $\mathcal{A}$* , such that:

- $(s_0, \mathbf{v}_0) \rightarrow_{\mathcal{A}}^{*\varepsilon} (s_{f(1)}, \mathbf{v}_{f(1)})$ , and
- for all  $i \geq 1$ ,  $(s_{f(i)}, \mathbf{v}_{f(i)}) \rightarrow_{\mathcal{A}}^{w[i]} (s_{f(i+1)}, \mathbf{v}_{f(i+1)})$  and  $(s_{f(i+1)}, \mathbf{v}_{f(i+1)}) \rightarrow_{\mathcal{A}}^{*\varepsilon} (s_{f(i+1)}, \mathbf{v}_{f(i+1)})$ .

A run  $\pi = (s_0, \mathbf{v}_0)(s_1, \mathbf{v}_1) \dots$  of  $w$  on  $\mathcal{A}$  is *accepting* iff  $|check_{\pi, k}^{\infty}| = +\infty$  for every  $k \in \{1, \dots, N\}$ . An  $\omega$ -word  $w \in \Sigma^{\omega}$  is *accepted* by  $\mathcal{A}$  iff there exists an accepting run of  $w$  on  $\mathcal{A}$ ; we denote by  $\mathcal{L}(\mathcal{A})$  the set of all  $\omega$ -words  $w \in \Sigma^{\omega}$  that are accepted by  $\mathcal{A}$ , and we say that  $\mathcal{A}$  *accepts* the language  $\mathcal{L}(\mathcal{A})$ . As an example, Figure 1 depicts a CCA with two counters ( $N = 2$ ) accepting the language  $((a^*b)^*a^Tb)^{\omega}$ . (Note that an automaton for the same language with one counter only can be devised as well.)

### 3.1 Decidability of the emptiness problem

We now prove that the emptiness problem for CCA is decidable in PTIME. The proof consists of 3 steps: (i) we replace general CCA by simple ones; (ii) we prove that their emptiness can be decided by checking the existence of finite witnesses of accepting runs; (iii) we show that the latter can be verified by checking for emptiness a suitable NFA.

**Simple CCA.** A CCA  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  is *simple* iff for each  $s \in S$  either  $|\{(s, \sigma, s', (k, op)) \in \Delta\}| = 1$  or  $op = no\_op$ ,  $k = 1$ , and  $\sigma = \varepsilon$  for all  $(s, \sigma, s', (k, op)) \in \Delta$ . Basically, a simple CCA has states of two kinds: those in which it can fire exactly one action and those in which it makes a nondeterministic choice. Moreover, for all pairs of configurations  $(s, \mathbf{v}), (s', \mathbf{v}')$  with  $(s, \mathbf{v}) \rightarrow_{\mathcal{A}}^{\sigma} (s', \mathbf{v}')$ , the transition  $\delta \in \Delta$  that has been fired in  $(s, \mathbf{v})$  is uniquely determined by  $s$  and  $s'$ . By exploiting  $\varepsilon$ -*transitions*, that is, transitions of the form  $(s, \varepsilon, s', (k, op))$ , and by adding a suitable number of states, it can be easily shown that every CCA  $\mathcal{A}$  may be turned into a simple one  $\mathcal{A}'$  such that  $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$ . Without loss of generality, in the rest of the section we restrict our attention to simple CCA.

The set of states of a CCA can be partitioned in four subsets: (i) the set of states  $s$  from which only one transition of the form  $(s, \sigma, s', (k, check))$  can be fired ( $check_k$  states); (ii) the set of states  $s$  from which only one transition of the form  $(s, \sigma, s', (k, inc))$  can be fired ( $inc_k$  states); (iii) the set of states  $s$  from which only one transition of the form  $(s, \sigma, s', (1, no\_op))$ , with  $\sigma \neq \varepsilon$ , can be fired ( $sym$  states); (iv) the set of states  $s$  from which possibly many transitions of the form  $(s, \varepsilon, s', (1, no\_op))$  can be fired ( $choice$  states).

Let  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  be a CCA. A *prefix computation* of  $\mathcal{A}$  is a finite prefix of a computation of  $\mathcal{A}$ ; formally, it is a finite sequence  $\mathcal{P} = (s_0, \mathbf{v}_0) \dots (s_n, \mathbf{v}_n)$  such that, for all  $i \in \{0, \dots, n-1\}$ ,  $(s_i, \mathbf{v}_i) \xrightarrow{\sigma_i}_{\mathcal{A}} (s_{i+1}, \mathbf{v}_{i+1})$ , for some  $\sigma_i \in \Sigma \cup \{\varepsilon\}$ . We denote by  $Prefixes_{\mathcal{A}}$  the sets of all prefix computations of  $\mathcal{A}$ . For every prefix computation  $\mathcal{P} = (s_0, \mathbf{v}_0) \dots (s_n, \mathbf{v}_n) \in Prefixes_{\mathcal{A}}$  and  $s \in S$ , it holds that if  $(s_n, \mathbf{v}_n) \xrightarrow{\sigma}_{\mathcal{A}} (s, \mathbf{v})$ , for some counter vector  $\mathbf{v}$  and some  $\sigma \in \Sigma \cup \{\varepsilon\}$ , then  $\mathbf{v}$  is uniquely determined by  $s_n, \mathbf{v}_n$ , and  $s$ , that is, there is no  $\mathbf{v}' \neq \mathbf{v}$  such that  $(s_n, \mathbf{v}_n) \xrightarrow{\sigma'}_{\mathcal{A}} (s, \mathbf{v}')$ , for any  $\sigma'$ .

**Finite witnesses of accepting runs.** We show now how to decide CCA emptiness by making use of the notion of accepting witness for a CCA.

**Definition 2** (Accepting witness). *Let  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  be a CCA. A prefix computation  $\mathcal{P} = (s_0, \mathbf{v}_0) \dots (s_n, \mathbf{v}_n) \in Prefixes_{\mathcal{A}}$  is an accepting witness (for  $\mathcal{A}$ ) iff there are  $2N + 2$  indexes  $begin < b_1 < e_1 < \dots < b_N < e_N < end$  such that  $0 \leq begin, end \leq n$ , and the following conditions hold:*

1. a non- $\varepsilon$ -transition can be fired from  $s_{begin}$ ;
2.  $s_{begin} = s_{end}$  and, for each  $k \in \{1, \dots, N\}$ ,  $s_{b_k} = s_{e_k}$ ,  $s_{b_k}$  is an  $inc_k$  state, and  $s_j$  is not a  $check_k$  state for any  $j$  with  $b_k \leq j \leq e_k$ ;
3. for each  $k \in \{1, \dots, N\}$ , there is  $j$ , with  $e_N < j < end$ , such that  $s_j$  is a  $check_k$  state.

An accepting witness for  $\mathcal{A}$  can be seen as a finite representation of an accepting run of some  $\omega$ -word on  $\mathcal{A}$ . Thus, deciding whether a CCA  $\mathcal{A}$  accepts the empty language amounts to searching  $Prefixes_{\mathcal{A}}$  for accepting witnesses. (The proof of the next lemma is omitted for lack of space.)

**Lemma 1.** *Let  $\mathcal{A}$  be a CCA. Then,  $\mathcal{L}(\mathcal{A}) \neq \emptyset$  iff  $Prefixes_{\mathcal{A}}$  contains an accepting witness.*

**From CCA to NFA.** Thanks to Lemma 1, deciding the emptiness problem for a CCA  $\mathcal{A}$  amounts to searching  $Prefixes_{\mathcal{A}}$  for an accepting witness. Since we restricted ourselves to simple CCA, we can safely identify elements of  $Prefixes_{\mathcal{A}}$  with their sequence of states and thus, by slightly abusing the notation, we can write, e.g.,  $s_0 s_1 \dots s_n \in Prefixes_{\mathcal{A}}$  for  $(s_0, \mathbf{v}_0) \dots (s_n, \mathbf{v}_n) \in Prefixes_{\mathcal{A}}$ . Given a CCA  $\mathcal{A}$ , let  $\mathcal{L}_w(\mathcal{A})$  be the language of finite words over the alphabet  $S$  (the set of states of  $\mathcal{A}$ ) that are accepting witnesses for  $\mathcal{A}$ . It is easy to see that  $\mathcal{L}(\mathcal{A}) \neq \emptyset$  if and only if  $\mathcal{L}_w(\mathcal{A}) \neq \emptyset$ . In what follows, for a CCA  $\mathcal{A}$  we build a nondeterministic finite automata (NFA) whose language is exactly  $\mathcal{L}_w(\mathcal{A})$ . Since the emptiness problem for NFA is decidable, so is the one for CCA.

In what follows, without loss of generality, we restrict our attention to accepting witnesses for which the set of indexes required by item 3 of Definition 2 is ordered. More precisely (we borrow the notation from Definition 2), we assume that there are  $N$  indexes  $c_1 < \dots < c_N$ , with  $e_N < c_1$  and  $c_N < end$ , such that  $s_{c_k}$  is a  $check_k$  state, for each  $k \in \{1, \dots, N\}$  (this requirement strengthens the one imposed by item 3 of Definition 2). Given a CCA  $\mathcal{A}$ , it is easy to check that  $Prefixes_{\mathcal{A}}$  contains an accepting witness, as specified by Definition 2, if and only if it contains one satisfying the additional ordering property above. Thus, Lemma 1 holds with respect to the new definition of accepting witness as well.

Given a CCA  $\mathcal{A}$ , we apply the following steps to build an NFA  $\mathcal{N}$  such that  $\mathcal{L}(\mathcal{N}) = \mathcal{L}_w(\mathcal{A})$ : (i) we build an NFA  $\mathcal{N}_1$  accepting finite words over the set of states of  $\mathcal{A}$  that are potential accepting witnesses, i.e., they satisfy conditions 1- 3 of Definition 2 but they might not be prefix computations; in other words, such an automaton might as well accept words not belonging to  $Prefixes_{\mathcal{A}}$ ; (ii) since



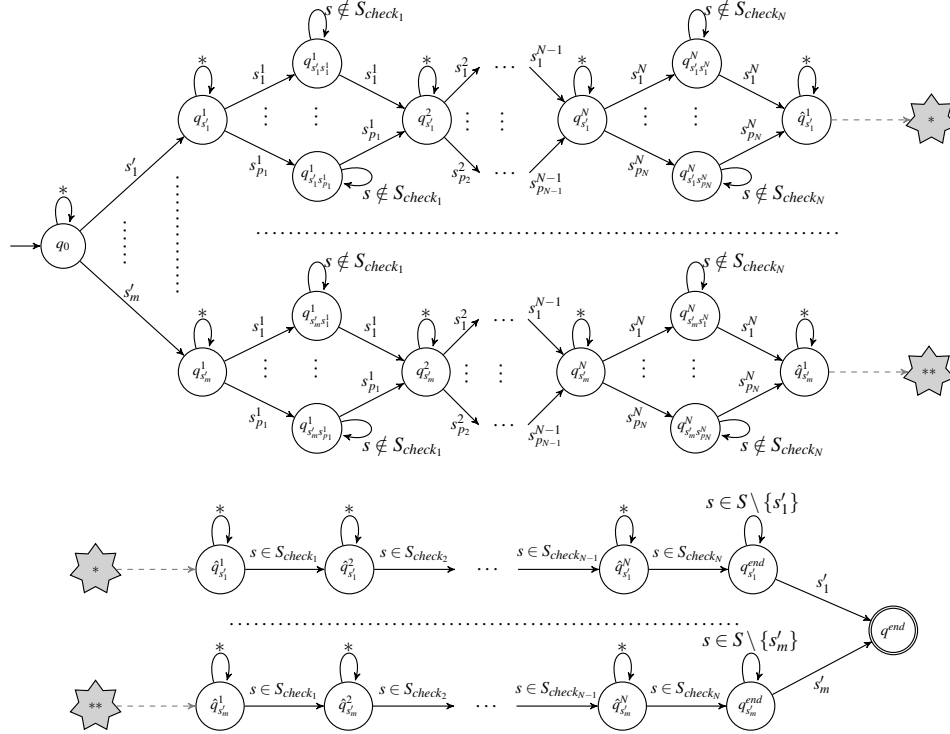


Figure 3: A graphical account of the automaton  $\mathcal{N}_1$ :  $S_{non-\varepsilon} = \{s'_1, s'_2, \dots, s'_m\}$ ,  $S_{inc_k} = \{s_1^k, s_2^k, \dots, s_{p_k}^k\}$  ( $k \in \{1, \dots, N\}$ ).

$Prefixes_{\mathcal{A}}$  is a regular language, thanks to closure properties of NFA, there exists an NFA  $\mathcal{N}$  whose language is  $\mathcal{L}(\mathcal{N}_1) \cap Prefixes_{\mathcal{A}} = \mathcal{L}_w(\mathcal{A})$ .

Let  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  be a CCA. We define  $\mathcal{N}_1 = \langle Q, \Sigma_{\mathcal{N}_1}, \delta, q_0, F \rangle$  as follows. We set  $\Sigma_{\mathcal{N}_1} = S$ ,  $F = \{q^{end}\}$ ; moreover, let  $S_{non-\varepsilon}$  be the set of states of  $S$  from which a non- $\varepsilon$ -transition can be fired, and  $S_{inc_k}$  be the sets of  $inc_k$  states in  $S$  ( $k \in \{1, \dots, N\}$ ), we set  $Q = \{q_0, q^{end}\} \cup \{q_{s'}^{end} \mid s' \in S_{non-\varepsilon}\} \cup \bigcup_{k=1}^N \{q_{s'}^k, \hat{q}_{s'}^k \mid s' \in S_{non-\varepsilon}\} \cup \bigcup_{k=1}^N \{q_{s', s''}^k \mid s' \in S_{non-\varepsilon}, s'' \in S_{inc_k}\}$ . The transition relation  $\delta$  is described in Figure 3. In particular, the automaton behaves as follows:

1. it nondeterministically guesses index *begin* when a symbol  $s' \in S_{non-\varepsilon}$  is read; the next state  $q_{s'}^1$  reached by  $\mathcal{N}_1$  stores the information about the state  $s'$  of  $\mathcal{A}$  being read to check, at a later stage (when index *end* is guessed), that  $s_{begin} = s' = s_{end}$ ;
2. similarly, for each  $k \in \{1, \dots, N\}$ , it nondeterministically guesses indexes  $b_k$  and  $e_k$ , when a symbol  $s^k$  corresponding to an  $inc_k$  state (of  $\mathcal{A}$ ) is read; once again, the information about the state  $s^k$  of  $\mathcal{A}$  being read is stored in the next state  $q_{s', s^k}^k$  reached by  $\mathcal{N}_1$ , in order to check that the same state  $s^k$  is read when  $e_k$  is guessed ( $s_{b_k} = s^k = s_{e_k}$ ); moreover, the automaton forces the absence of  $check_k$  state in between indexes  $b_k$  and  $e_k$ ;
3. it checks for the existence, after  $e_N$ , of  $check_k$  states ( $k \in \{1, \dots, N\}$ ) in the desired order;
4. wait for the input symbol  $s'$ , that is, the same symbol read when *begin* was guessed; when such a symbol is read,  $\mathcal{N}_1$  enters the final state  $q^{end}$ .

Let  $S_{non-\varepsilon}$  and  $S_{inc_k}$  ( $k \in \{1, \dots, N\}$ ) be defined as above and, in addition, let  $S_{check_k}$  be the set of  $check_k$  states in  $S$ . We formally define  $\delta$  as follows:

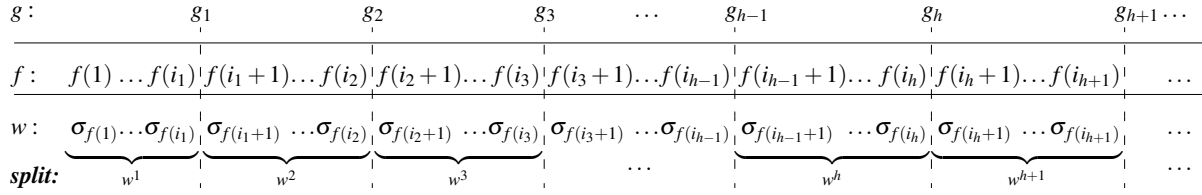


Figure 4: An infinite word  $w = w[1]w[2] \dots w[i] \dots = \sigma_{f(1)}\sigma_{f(2)} \dots \sigma_{f(i)} \dots$  is split using sequence  $g_1 < g_2 < \dots < g_h < \dots$  into infinitely many finite words  $w^1, w^2, \dots, w^h, \dots$ .

$$\begin{aligned} \delta = & \{(q_0, s, q_0) \mid s \in S\} \cup \{(q_0, s', q_{s'}^1) \mid s' \in S_{non-\varepsilon}\} \cup \bigcup_{k=1}^N \{(q_{s'}^k, s, q_{s'}^k) \mid s' \in S_{non-\varepsilon}, s \in S\} \\ & \cup \bigcup_{k=1}^N \{(q_{s'}^k, s^k, q_{s',s^k}^k) \mid s' \in S_{non-\varepsilon}, s^k \in S_{inc_k}\} \cup \bigcup_{k=1}^N \{(q_{s',s^k}^k, s, q_{s',s^k}^k) \mid s' \in S_{non-\varepsilon}, s^k \in S_{inc_k}, s \in S \setminus S_{check_k}\} \\ & \cup \bigcup_{k=1}^{N-1} \{(q_{s',s^k}^k, s^k, q_{s',s^{k+1}}^{k+1}) \mid s' \in S_{non-\varepsilon}, s^k \in S_{inc_k}\} \cup \{(q_{s',s^N}^N, s^N, \hat{q}_{s'}^1) \mid s' \in S_{non-\varepsilon}, s^N \in S_{inc_N}\} \\ & \cup \bigcup_{k=1}^N \{(q_{s'}^k, s, q_{s'}^k) \mid s' \in S_{non-\varepsilon}, s \in S\} \cup \bigcup_{k=1}^{N-1} \{(q_{s'}^k, s^k, \hat{q}_{s'}^{k+1}) \mid s' \in S_{non-\varepsilon}, s^k \in S_{check_k}\} \\ & \cup \{(\hat{q}_{s'}^N, s^N, q_{s'}^{end}) \mid s' \in S_{non-\varepsilon}, s^N \in S_{check_N}\} \cup \{(q_{s'}^{end}, s, q_{s'}^{end}) \mid s' \in S_{non-\varepsilon}, s \in S \setminus \{s'\}\} \\ & \cup \{(q_{s'}^{end}, s', q_{s'}^{end}) \mid s' \in S_{non-\varepsilon}\} \end{aligned}$$

Since the size of  $\mathcal{N}_1$  is polynomial in the size of  $\mathcal{A}$  ( $|\mathcal{Q}| \leq 2 + 2 \cdot N \cdot |S| + N \cdot |S|^2 + |S|$ ), we have a polynomial reduction from the emptiness problem for CCA to the one for NFA.

**Theorem 1.** *The emptiness problem for CCA is decidable in PTIME.*

## 4 From $\omega$ T-regular languages to CCA

In this section, we show how to map an  $\omega$ T-regular expression  $E$  into a corresponding CCA  $\mathcal{A}$  such that  $\mathcal{L}(E) = \mathcal{L}(\mathcal{A})$ . We build the automaton  $\mathcal{A}$  in a compositional way: for each sub-expression  $E'$  of  $E$ , starting from the atomic ones, we introduce a set  $\mathcal{S}_{E'}$  of CCAs and then we show how to produce the set of automata for complex sub-expressions by suitably combining automata in the sets associated with their sub-expressions. Eventually, we obtain a set of automata for the  $\omega$ T-regular expression  $E$ . The automaton  $\mathcal{A}$  results from the merge of the automata in such a set, as described below. Without loss of generality, we assume the sets of states of all automata generated in the construction to be pairwise disjoint, i.e., if  $\mathcal{A}' \in \mathcal{S}_{E'}$  and  $\mathcal{A}'' \in \mathcal{S}_{E''}$ , where  $E'$  and  $E''$  are two (not necessarily distinct) sub-expressions of  $E$ , then the set of states of  $\mathcal{A}'$  and the one of  $\mathcal{A}''$  are disjoint.

We proceed by structural induction on  $\omega$ T-regular expressions, that is, when building the set  $\mathcal{S}_{E'}$  of CCAs for a sub-expression  $E'$  of  $E$ , we assume the sets of CCAs for the sub-expressions of  $E'$  to be available. In addition, by construction, we force each generated CCA  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  to feature a distinguished *final state*  $s_f$  such that  $(s_f, \sigma, s', (k, op)) \in \Delta$  implies  $\sigma = \varepsilon$ ,  $s' = s_f$ ,  $k = 1$ , and  $op = inc$ ; in order to distinguish the final state of a CCA we sometimes abuse the notation and write  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta)$ , where  $s_f$  is the final state of  $\mathcal{A}$ .

**Encoding of T-regular expressions.** We first deal with T-regular expressions (sub-grammar rooted in  $e$  in paragraph “ $\omega$ T-regular languages” at page 5). Since a T-regular expression produces a language of word sequences and our automata accept  $\omega$ -words, we must find a way to extract sequences from  $\omega$ -words. Intuitively, we do that by splitting an infinite word into infinitely many finite sub-words, each of them corresponding to the sequence of symbols in between two consecutive *check* of the 1st counter along the corresponding accepting run. Formally, let  $\pi = (s_0, \mathbf{v}_0)(s_1, \mathbf{v}_1) \dots$  be an accepting run of some  $\omega$ -word  $w$  on some CCA  $\mathcal{A}$  such that  $(s_i, \mathbf{v}_i) \xrightarrow{\sigma_i}_{\mathcal{A}} (s_{i+1}, \mathbf{v}_{i+1})$  via  $\delta_i$ , for each  $i \geq 0$ , and let  $f$  be the trace of  $w$  in  $\pi$  wrt.  $\mathcal{A}$  (see definition of run at page 7). Recall that  $f$  is such that  $\sigma_{f(i)} = w[i]$  for all

$i \geq 1$  (roughly speaking,  $f$  enumerates symbols different from  $\varepsilon$  within sequence  $\sigma_0\sigma_1\dots$ ). Moreover, let  $g_1 < g_2 < \dots < g_h < \dots$  ( $g_h \in \mathbb{N}$  for every  $h$ ) be the sequence of indexes corresponding to transitions in  $\pi$  where the 1st counter is checked, that is, for every  $i \in \mathbb{N}$  we have that  $\delta_i$  has the form  $(s_i, \sigma_i, s_{i+1}, (1, \text{check}))$  if and only if  $i = g_h$  for some  $h$ . As shown in Figure 4, the sequence  $\langle g_h \rangle_{h \in \mathbb{N}_{>0}}$  defines a unique partition of the infinite word  $w = \sigma_{f(1)}\sigma_{f(2)}\sigma_{f(3)}\dots$  into infinitely many finite sub-words (some of them are possibly empty words):  $w^1 = \sigma_{f(1)}\sigma_{f(2)}\dots\sigma_{f(i_h)}$ ,  $w^2 = \sigma_{f(i_h+1)}\sigma_{f(i_h+2)}\dots\sigma_{f(i_{h+1})}$ ,  $w^3 = \sigma_{f(i_{h+1}+1)}\dots\sigma_{f(i_{h+2})}$ ,  $\dots$ ,  $w^h = \sigma_{f(i_{h-1}+1)}\dots\sigma_{f(i_h)}$ , and so on, with  $f(i_h) < g_h \leq f(i_h + 1)$  for every  $h$ . We define the language of word sequences accepted by  $\mathcal{A}$ , denoted by  $\mathcal{L}_s(\mathcal{A})$ , as  $\mathcal{L}_s(\mathcal{A}) = \{(w^1, w^2, \dots, w^h, \dots) : w \in \mathcal{L}(\mathcal{A})\}$ .

Let  $\widehat{\mathcal{A}} = (S, \Sigma, s_0, s_f, N, \Delta \cup \{(s_f, \varepsilon, s_0, (1, \text{check}))\})$ , for every  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta)$ . For each expression  $e$ , we build a set  $\mathcal{S}_e$  for which it holds:

$$\mathcal{L}(e) = \bigcup_{\mathcal{A} \in \mathcal{S}_e} \mathcal{L}_s(\widehat{\mathcal{A}}).$$

*Base cases.* If  $e = \emptyset$ , then  $\mathcal{S}_e = \{\mathcal{A}_\emptyset\}$  where  $\mathcal{A}_\emptyset = (\{s_0, s_f\}, \Sigma, s_0, s_f, 1, \emptyset)$ .

If  $e = a$ , then  $\mathcal{S}_e = \{\mathcal{A}_a\}$  where  $\mathcal{A}_a = (\{s_0, s_f\}, \Sigma, s_0, s_f, 1, \{(s_0, a, s_f, (1, \text{no\_op})), (s_f, \varepsilon, s_f, (1, \text{inc}))\})$ .

See Figure 5 (a) and (b) for a graphical account of both cases.

*Inductive step.* For our purposes, we define, for every CCA  $\mathcal{A} = (S, \Sigma, s_0, N, \Delta)$  and natural number  $N' \geq 1$ , the  $N'$ -shifted version of  $\mathcal{A}$  as the automaton  $\mathcal{A}' = (S, \Sigma, s_0, N + N', \{(s, \sigma, s, (k + N', \text{op})) : (s, \sigma, s, (k, \text{op})) \in \Delta\})$ . Four cases must be considered.

- Let  $e = e_1 \cdot e_2$ ,  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta) \in \mathcal{S}_{e_1}$ , and  $\mathcal{A}' = (S', \Sigma, s'_0, s'_f, N', \Delta') \in \mathcal{S}_{e_2}$ . Moreover, let  $\mathcal{A}'' = (S, \Sigma, s_0, s_f, N + 1, \Delta'')$  and  $\mathcal{A}''' = (S', \Sigma, s'_0, s'_f, N' + N + 1, \Delta''')$  be the 1-shifted version of  $\mathcal{A}$  and the  $N + 1$ -shifted version of  $\mathcal{A}'$ , respectively. We define  $\mathcal{A} \cdot \mathcal{A}' = (S \cup S' \cup \{s''_f\}, \Sigma, s_0, s''_f, N + N' + 1, \Delta'' \cup \Delta''' \cup \{(s_f, \varepsilon, s'_0, (2, \text{check})), (s'_f, \varepsilon, s''_f, (N + 2, \text{check})), (s''_f, \varepsilon, s''_f, (1, \text{inc}))\})$ .

We set  $\mathcal{S}_{e_1 \cdot e_2} = \{\mathcal{A} \cdot \mathcal{A}' : \mathcal{A} \in \mathcal{S}_{e_1}, \mathcal{A}' \in \mathcal{S}_{e_2}\}$ . See Figure 5 (c) for a graphical account.

- Let  $e = e_1 + e_2$ ,  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta) \in \mathcal{S}_{e_1}$ , and  $\mathcal{A}' = (S', \Sigma, s'_0, s'_f, N', \Delta') \in \mathcal{S}_{e_2}$ . Moreover, let  $\mathcal{A}''$  and  $\mathcal{A}'''$  be defined as in the previous case. We define  $\mathcal{A} + \mathcal{A}'$  as the set  $\{\mathcal{A}_{+1}, \mathcal{A}_{+2}, \mathcal{A}_{+3}\}$  (see Figure 5 (d)), where

- $\mathcal{A}_{+1} = (S \cup S' \cup \{\bar{s}_{01}, \bar{s}_{f1}\}, \Sigma, \bar{s}_{01}, \bar{s}_{f1}, N' + N + 1, \Delta'' \cup \Delta''' \cup \{(\bar{s}_{01}, \varepsilon, s_0, (1, \text{no\_op})), (\bar{s}_{01}, \varepsilon, s'_0, (1, \text{no\_op})), (s_f, \varepsilon, \bar{s}_{f1}, (2, \text{check})), (s'_f, \varepsilon, \bar{s}_{f1}, (N + 2, \text{check})), (\bar{s}_{f1}, \varepsilon, \bar{s}_{f1}, (1, \text{inc}))\} \cup \{(s_f, \varepsilon, s_f, (k, *)) : * \in \{\text{inc}, \text{check}\}, N + 2 \leq k \leq N + N' + 1\})$ ,
- $\mathcal{A}_{+2} = (S \cup S' \cup \{\bar{s}_{02}, \bar{s}_{f2}\}, \Sigma, \bar{s}_{02}, \bar{s}_{f2}, N' + N + 1, \Delta'' \cup \Delta''' \cup \{(\bar{s}_{02}, \varepsilon, s_0, (1, \text{no\_op})), (\bar{s}_{02}, \varepsilon, s'_0, (1, \text{no\_op})), (s_f, \varepsilon, \bar{s}_{f2}, (2, \text{check})), (s'_f, \varepsilon, \bar{s}_{f2}, (N + 2, \text{check})), (\bar{s}_{f2}, \varepsilon, \bar{s}_{f2}, (1, \text{inc}))\} \cup \{(s'_f, \varepsilon, s'_f, (k, *)) : * \in \{\text{inc}, \text{check}\}, 2 \leq k \leq N + 1\})$ , and
- $\mathcal{A}_{+3} = (S \cup S' \cup \{\bar{s}_{03}, \bar{s}_{f3}\}, \Sigma, \bar{s}_{03}, \bar{s}_{f3}, N' + N + 1, \Delta'' \cup \Delta''' \cup \{(\bar{s}_{03}, \varepsilon, s_0, (1, \text{no\_op})), (\bar{s}_{03}, \varepsilon, s'_0, (1, \text{no\_op})), (s_f, \varepsilon, \bar{s}_{f3}, (2, \text{check})), (s'_f, \varepsilon, \bar{s}_{f3}, (N + 2, \text{check})), (\bar{s}_{f3}, \varepsilon, \bar{s}_{f3}, (1, \text{inc}))\} \cup \{(s_f, \varepsilon, s_f, (k, *)) : * \in \{\text{inc}, \text{check}\}, 2 \leq k \leq N + 1\} \cup \{(s'_f, \varepsilon, s'_f, (k, *)) : * \in \{\text{inc}, \text{check}\}, N + 2 \leq k \leq N + N' + 1\})$ .

We set  $\mathcal{S}_{e_1 + e_2} = \bigcup_{\mathcal{A} \in \mathcal{S}_{e_1}, \mathcal{A}' \in \mathcal{S}_{e_2}} \mathcal{A} + \mathcal{A}'$ .

- Let  $e = e_1^*$ ,  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta) \in \mathcal{S}_{e_1}$ , and  $\mathcal{A}''$  be defined as in the previous cases. We let  $\mathcal{A}_* =$

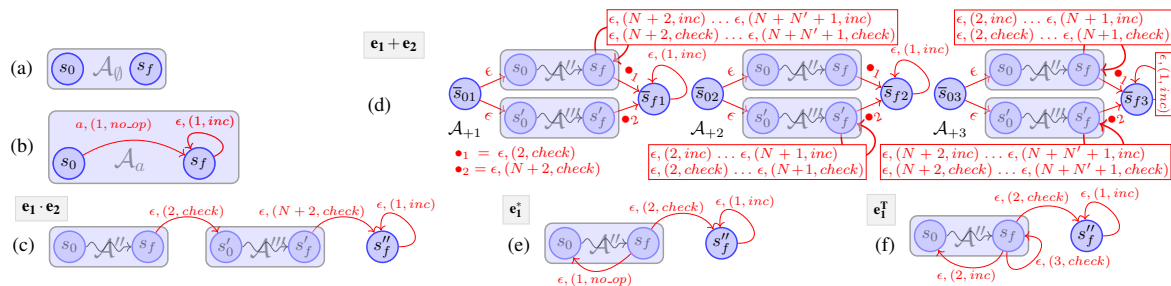


Figure 5: The automata for the translation of a  $T$ -regular expression  $e$ .

$(S \cup \{s''_f\}, \Sigma, s_0, s''_f, N+1, \Delta'' \cup \{(s_f, \varepsilon, s_0, (1, no\_op)), (s_f, \varepsilon, s''_f, (2, check)), (s''_f, \varepsilon, s''_f, (1, inc))\})$ .

We set  $\mathcal{S}_{e_1^*} = \{\mathcal{A}_* : \mathcal{A} \in \mathcal{S}_{e_1}\}$ . See Figure 5 (e) for a graphical account.

- Let  $e = e_1^T$  and  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta) \in \mathcal{S}_{e_1}$ . Moreover, let  $\mathcal{A}'' = (S, \Sigma, s_0, s_f, N+2, \Delta'')$  be the 2-shifted version of  $\mathcal{A}$ . We let  $\mathcal{A}_T = (S \cup \{s''_f\}, s_0, s''_f, N+2, \Delta'' \cup \{(s_f, \varepsilon, s_0, (2, inc)), (s_f, \varepsilon, s''_f, (2, check)), (s''_f, \varepsilon, s''_f, (3, check)), (s''_f, \varepsilon, s''_f, (1, inc))\})$ .

We set  $\mathcal{S}_{e_1^T} = \{\mathcal{A}_T : \mathcal{A} \in \mathcal{S}_{e_1}\}$ . See Figure 5 (f) for a graphical account.

The next lemma states the correctness of the proposed encoding (proof omitted for lack of space).

**Lemma 2.** *Let  $e$  be a  $T$ -regular expression and  $\mathcal{S}_e$  be the corresponding set of automata. It holds:*

$$\mathcal{L}(e) = \bigcup_{\mathcal{A} \in \mathcal{S}_e} \mathcal{L}_s(\widehat{\mathcal{A}}).$$

**Encoding of  $\omega T$ -regular expressions.** We are now ready to deal with  $\omega T$ -regular expressions (sub-grammar rooted in  $E$  in paragraph “ $\omega T$ -regular languages” at page 5). We must distinguish three cases.

- If  $E = E_1 + E_2$ , then  $\mathcal{S}_{E_1+E_2}$  is equal to  $\mathcal{S}_{E_1} \cup \mathcal{S}_{E_2}$ .
- If  $E = R \cdot E'$ , then let  $A_R = (S_R, F_R, \Sigma, s_0^R, \Delta_R)$  be the NFA that recognises the regular language  $\mathcal{L}(R)$ , and  $\mathcal{A} = (S, \Sigma, s_0, s_f, N, \Delta) \in \mathcal{S}_{E'}$ . We let  $A_R \cdot \mathcal{A} = (S \cup S_R, \Sigma, s_0^R, s_f, N, \Delta \cup \{(s, \sigma, s', (1, no\_op)) : (s, \sigma, s') \in \Delta_R\} \cup \{(s, \varepsilon, s_0, (1, no\_op)) : s \in F_R\})$ . We set  $\mathcal{S}_{R \cdot E'} = \{A_R \cdot \mathcal{A} : \mathcal{A} \in \mathcal{S}_{E'}\}$ .
- Finally, if  $E = e^\omega$ , then  $\mathcal{S}_{e^\omega}$  is the set  $\{\widehat{\mathcal{A}} : \mathcal{A} \in \mathcal{S}_e\}$ .

As in the case of  $T$ -regular expressions, it is easy to check that, for all  $\omega T$ -regular expressions  $E$ :

$$\mathcal{L}(E) = \bigcup_{\mathcal{A} \in \mathcal{S}_E} \mathcal{L}(\widehat{\mathcal{A}}).$$

To complete the reduction, we only need to show how to merge the automata in  $\mathcal{S}_E$  into a single one  $\mathcal{A}_E$  accepting the language  $\mathcal{L}(E)$ . Let  $\mathcal{S}_E = \{\mathcal{A}_1, \dots, \mathcal{A}_n\}$ , with  $\mathcal{A}_i = (S_i, \Sigma, s_0^i, N_i, \Delta_i)$ , for  $1 \leq i \leq n$ , and let  $N_{\max} = \max\{N_i : 1 \leq i \leq n\}$ . For each  $1 \leq i \leq n$ , let  $\bar{\Delta}_i = \Delta_i \cup \{(s_0^i, \varepsilon, s_0^i, (k, *)) : * \in \{inc, check\}, N_i < k \leq N_{\max}\}$ . Finally, let  $s_0$  be a fresh state. We define  $\mathcal{A}_E$  as the automaton  $(\bigcup_{1 \leq i \leq n} S_i \cup \{s_0\}, \Sigma, s_0, N_{\max}, \bigcup_{1 \leq i \leq n} (\bar{\Delta}_i \cup \{(s_0, \varepsilon, s_0^i, (1, no\_op))\}))$ .

**Theorem 2.** *For every  $\omega T$ -regular expression  $E$ , there exists a CCA  $\mathcal{A}$  such that  $\mathcal{L}(E) = \mathcal{L}(\mathcal{A})$ .*

## 5 From $\omega T$ -regular languages to S1S+U

In this section, we provide an encoding of  $\omega T$ -regular expressions into S1S+U.

**Definition of S1S+U.** The logic S1S is MSO interpreted over infinite words. Its formulas are built over a finite, non-empty alphabet  $\Sigma$  and sets  $V_1$  and  $V_2$  of first- and second-order variables, respectively:

$$\begin{aligned} \varphi &::= \tau \in P_\sigma \mid \tau \in X \mid \neg \varphi \mid \varphi \vee \varphi \mid \exists x. \varphi \mid \exists X. \varphi \\ \tau &::= x \mid s(\tau) \end{aligned}$$

where  $\sigma \in \Sigma$ ,  $x \in V_1$ , and  $X \in V_2$ . We denote by  $V_\Sigma$  the set  $\{P_\sigma \mid \sigma \in \Sigma\}$ .<sup>5</sup> Technically, elements of  $V_\Sigma$  are second-order variables (i.e., they range over sets of positive natural numbers), but with a standard intended semantics: they partition  $\mathbb{N}_{>0}$  and an interpretation for them  $\mathcal{I} : V_\Sigma \rightarrow 2^{\mathbb{N}_{>0}}$  identifies an infinite word  $w^\mathcal{I}$  over  $\Sigma$  as follows:  $w^\mathcal{I}[i] = \sigma$  iff  $i \in \mathcal{I}(P_\sigma)$ , for every  $i \in \mathbb{N}_{>0}$ ,  $\sigma \in \Sigma$ . Notice also that variables in  $V_\Sigma$  always occur free (i.e., not bound by any quantifier). A formula is *closed* if the only free variables are the ones in  $V_\Sigma$ ; otherwise, it is *open*. The semantics of a closed formula  $\varphi$ , denoted by  $\llbracket \varphi \rrbracket$ , is the set of all infinite words that *satisfy*  $\varphi$ , i.e.,  $\llbracket \varphi \rrbracket = \{w^\mathcal{I} \mid \mathcal{I} \models \varphi\}$ .

The logic S1S+U extends S1S with the *unbounding quantifier*  $\mathbb{U}$ , which is defined as in [4]:

$$\mathbb{U}X. \varphi(X) := \bigwedge_{n \in \mathbb{N}} \exists_{\text{fin}} X (\varphi(X) \wedge |X| \geq n).$$

where  $\exists_{\text{fin}}$  allows for existential quantification over finite sets, i.e.,  $\exists_{\text{fin}} X. \varphi \equiv \exists X. (\varphi \wedge \exists y. X \subseteq \{1, \dots, y\})$  for every second-order variable  $X$  and S1S+U-formula  $\varphi$ ; the universal quantifier  $\forall_{\text{fin}}$  is defined as the dual

<sup>5</sup>We also use the formulation  $P(\tau)$  in stead of  $\tau \in P$  ( $P \in V_2 \cup V_\Sigma$ ).

of  $\exists_{\text{fin}}$ . Intuitively,  $\mathbb{U}$  makes it possible to say that a formula  $\varphi(X)$  (containing at least one second-order free variable  $X$ ) is satisfied by infinitely many finite sets and there is no bound on their sizes. The *bounding quantifier*  $\mathbb{B}$  is defined as the negation of  $\mathbb{U}$ :  $\mathbb{B}X.\varphi(X) := \neg\mathbb{U}X.\varphi(X) \equiv \forall_{n \in \mathbb{N}} \forall_{\text{fin}} X (\varphi(X) \rightarrow |X| < n)$ . Its intended meaning is: there is a bound on the sizes of finite sets that satisfy  $\varphi(X)$ .

**Encoding.** In what follows, given an  $\omega T$ -regular expression  $E$  we show how to build a formula  $\varphi_E$  for which  $\mathcal{L}(E) = \llbracket \varphi_E \rrbracket$ . For the lack of space, we only give an intuitive idea of the encoding.

For every  $\omega T$ -regular (sub-)expression  $E$ , let  $E_{[T \mapsto *]}$  be the  $\omega$ -regular (sub-)expression obtained from  $E$  by replacing the  $T$ -constructor with the  $*$ -constructor (e.g., if  $E = (a^T b)^\omega$ , then  $E_{[T \mapsto *]} = (a^* b)^\omega$ ) and  $\varphi_{E_{[T \mapsto *]}}$  be the S1S-formula for which  $\llbracket \varphi_{E_{[T \mapsto *]}} \rrbracket = \mathcal{L}(E_{[T \mapsto *]})$  holds (its existence is guaranteed by the equivalence between S1S and  $\omega$ -regular languages).

Let  $E$  be an  $\omega T$ -regular expression. In order to correctly define  $\varphi_E$  we need to enrich such a formula  $\varphi_{E_{[T \mapsto *]}}$  to enforce the condition imposed by occurrences of the  $T$ -constructor in  $E$ . The intuitive idea is to control, for every sub-expression  $e^T$ , the sizes of  $e$ -blocks (i.e., maximal blocks of consecutive occurrences of finite words in  $\mathcal{L}(e_{[T \mapsto *]})$ ) along infinite words. (Notice that  $e_{[T \mapsto *]}$  is a regular expression.) According to the semantics of the  $T$ -constructor, we have to force the existence of  $e$ -blocks of infinitely many different sizes, and infinitely many of such sizes must occur infinitely often. To this end, given a regular expression  $e$ , we build a formula  $\Phi_e^{T\text{cond}}$  that is satisfied by an infinite word  $w$  iff there are infinitely many  $k \in \mathbb{N}$  such that  $w$  features infinitely many  $e$ -blocks of size  $k$ . In our construction, we use formulas  $is\_reg\_exp_e(x, y)$  (for every regular expression  $e$ ), featuring two free first-order variables, with the following semantics:  $w$  satisfies  $is\_reg\_exp_e[x \mapsto \bar{x}, y \mapsto \bar{y}]$  iff  $w[\bar{x}, \bar{y}] \in \mathcal{L}(e)$ . In addition, we use the unary predicate  $Beginning\_of_e(x)$ , with the following semantics:  $w$  satisfies  $Beginning\_of_e[x \mapsto \bar{x}]$  iff  $w[\bar{x}, \bar{y}] \in \mathcal{L}(e)$  for some  $\bar{y} \in \mathbb{N}_{>0}$ .

Let  $e$  be a regular expression. To begin with, we define formula  $\Phi_{e\text{-block}}(X)$ , stating that  $X$  is a maximal set of positions from which consecutive sub-words belonging to  $\mathcal{L}(e)$  begin; roughly speaking,  $X$  is an  $e$ -block.

$$\Phi_{e\text{-block}}(X) := \exists y \exists z. [is\_reg\_exp_{e^*}(y, z) \wedge X \subseteq \{y, \dots, z\} \wedge \forall x. (x \in \{y, \dots, z\} \wedge Beginning\_of_e(x) \rightarrow x \in X)].$$

Next formula  $\Phi_{e\text{-block-set}}(Y)$  says that (i)  $Y$  only contains  $e$ -blocks, (ii) it contains infinitely many of them, and (iii) there is an upper bound on their sizes. In this case, we say that  $Y$  is an  $e$ -block-set.

$$\Phi_{e\text{-block-set}}(Y) := [\forall y. (y \in Y \rightarrow \exists_{\text{fin}} X. (\Phi_{e\text{-block}}(X) \wedge X \subseteq Y \wedge y \in X))] \wedge \\ \wedge [\forall y \exists_{\text{fin}} X. (\Phi_{e\text{-block}}(X) \wedge X \subseteq Y \wedge \min X > y)] \wedge [\mathbb{B}x. (X \subseteq Y \wedge \Phi_{e\text{-block}}(X))]$$

Finally, we define  $\Phi_e^{T\text{cond}}$  as  $\forall Y. [\Phi_{e\text{-block-set}}(Y) \rightarrow \exists Z. (\Phi_{e\text{-block-set}}(Z) \wedge Y \subsetneq Z \wedge \exists^\omega x. x \in Z \setminus Y)]$ , where  $\exists^\omega$  allows for infinite existential first-order quantification, i.e.,  $\exists^\omega x. \varphi \equiv \forall y \exists x. (x > y \wedge \varphi)$ .

**Lemma 3.** *Let  $e$  be a regular expression. An infinite word  $w$  satisfies  $\Phi_e^{T\text{cond}}$  iff there are infinitely many natural numbers  $k$  such that  $w$  features infinitely many  $e$ -blocks of size  $k$ .*

*Proof.* Let  $w$  be an infinite word that satisfies  $\Phi_e^{T\text{cond}}$  and let us assume that there are only finitely many natural numbers  $k$  such that infinitely many  $e$ -blocks of size  $k$  occur in  $w$ . Let  $k_{\text{max}}$  be the largest among such numbers and let  $\bar{Y}$  be the  $e$ -block-set containing all  $e$ -blocks of size not larger than  $k_{\text{max}}$ . Clearly,  $w$  satisfies  $\Phi_{e\text{-block-set}}[Y \mapsto \bar{Y}]$  and thus, by the definition of  $\Phi_e^{T\text{cond}}$ , there exists an  $e$ -block-set  $\bar{Z}$  that contains infinitely many  $e$ -blocks (of bounded size) that do not belong to  $\bar{Y}$ . Since  $\bar{Z} \supsetneq \bar{Y}$  (that means  $Z$  contains all  $e$ -blocks in  $\bar{Y}$  as well), there exists a number  $k' > k_{\text{max}}$  such that infinitely many  $e$ -blocks of size  $k'$  occur in  $w$ . This is in contradiction with our initial hypothesis that  $k_{\text{max}}$  is the largest number such that infinitely many  $e$ -blocks of size  $k_{\text{max}}$  occur in  $w$ , hence the thesis follows.

In order to prove the converse direction, let us assume that there are infinitely many natural numbers  $k$  such that infinitely many  $e$ -blocks of size  $k$  occur in  $w$  and let  $\bar{Y}$  be an  $e$ -block-set (i.e.,  $w$  satisfies

$\Phi_{e\text{-block-set}}[Y \mapsto \bar{Y}]$ ). By the definition of  $\Phi_{e\text{-block-set}}(Y)$  (in particular, the third conjunct), there is a bound on the size of all  $e$ -blocks in  $\bar{Y}$ . Let  $k_{\max}$  be such a bound. By our assumption, there is a number  $k' > k_{\max}$  such that infinitely many  $e$ -blocks of size  $k'$  occur in  $w$ . Let  $\bar{Z}$  be the set containing all  $e$ -blocks in  $\bar{Y}$  and, in addition all  $e$ -blocks of size  $k'$  occurring in  $w$ . Clearly,  $\bar{Z}$  is an  $e$ -block-set that contains  $\bar{Y}$  and features infinitely many elements not belonging to  $\bar{Y}$  (i.e.,  $w$  satisfies the formula  $(\Phi_{e\text{-block-set}}(Z) \wedge Y \subsetneq Z \wedge \exists^\omega x.x \in Z \setminus Y)[Y \mapsto \bar{Y}, Z \mapsto \bar{Z}]$ ), and thus  $w$  satisfies  $\Phi_e^{T\text{cond}}$ .  $\square$

Making use of formulas  $\Phi_e^{T\text{cond}}$ , for every regular expression  $e$ , it is possible to strengthen  $\varphi_{E[T \mapsto *]}$  to enforce the condition, imposed by occurrences of the  $T$ -constructor in  $E$ , on sizes of  $e$ -blocks occurring in infinite words, for every sub-expression  $e^T$  of  $E$ . Thus, we can conclude the main result of this section.

**Theorem 3.** *For every  $\omega$ T-regular expression  $E$ , we have that  $\llbracket \varphi_E \rrbracket = \mathcal{L}(E)$ .*

As a conclusive remark, notice that  $\Phi_e^{T\text{cond}}$  uses quantification over infinite sets, implying that  $\varphi_t$  does not belong to the language of  $w$ S1S+U, where second-order quantification is only allowed over finite sets.

## 6 Conclusions

In this paper, we introduced a new class of extended  $\omega$ -regular languages ( $\omega$ T-regular languages), that captures meaningful languages not belonging to the class of  $\omega$ BS-regular ones. We first gave a characterization of them in terms of  $\omega$ T-regular expressions. Then, we defined the new class of counter-check automata (CCA), with a decidable emptiness problem, and we proved that they are expressive enough to capture them. Finally, we provided an embedding of  $\omega$ T-regular languages in S1S+U.

In the exploration of the space of possible extensions of  $\omega$ -regular languages, we studied also a stronger variant of  $(\cdot)^T$ , that forces  $\omega$ -words to feature infinitely many exponents, *all of them* occurring infinitely often (a detailed account can be found in [2]). To a large extent, the results obtained for  $(\cdot)^T$  can be replicated for this stronger variant. In particular, it is possible to introduce a new class of automata, called *counter-queue automata* (CQA), that generalize CCA, whose emptiness problem can be proved to be decidable in 2ETIME and which are expressive enough to capture  $\omega$ -regular languages extended with the stronger variant of  $(\cdot)^T$ . As in the case of  $\omega$ T-regular languages, the problem of establishing whether or not the new languages are expressively complete with respect to CQA is open. There are, however, at least two significant differences between  $(\cdot)^T$  and its stronger variant. First,  $(\cdot)^T$  satisfies the following property of *prefix independence*. Let  $e$  be a  $T$ -regular expression and let  $\vec{u} = (u_1, u_2, \dots)$  and  $\vec{v} = (u_h, u_{h+1}, \dots)$  be two word sequences such that  $\vec{v}$  is the infinite suffix of  $\vec{u}$  starting at position  $h$  and  $u_i \in \mathcal{L}(e)$  for all  $i$ . Then,  $\vec{u} \in \mathcal{L}(e^T)$  iff  $\vec{v} \in \mathcal{L}(e^T)$ . Both  $(\cdot)^B$  and  $(\cdot)^S$  satisfy an analogous property, while this is not the case with the stronger variant of  $(\cdot)^T$ : if  $\vec{u}$  belongs to the language, then  $\vec{v}$  belongs to it as well, but not vice versa. The second difference is that there seems to be no way to generalize the embedding of  $\omega$ T-regular languages into S1S+U given in Section 5 to the stronger variant of  $(\cdot)^T$ .

As for future work, we would like to investigate different combinations of  $(\cdot)^B$ ,  $(\cdot)^S$ , and (weak and strong)  $(\cdot)^T$ . We already know that  $\omega$ BST-regular languages are not closed under complementation. Indeed, if they were, they would be expressively complete for S1S+U. However, it is known from [10] that S1S+U makes it possible to define languages that are complete for arbitrary levels of the projective hierarchy, while  $\omega$ BST-regular languages live at the first level (analytic sets), and thus they cannot define full S1S+U. A particularly interesting issue is the one about the intersections of  $\omega$ B-,  $\omega$ S-, and weak/strong  $\omega$ B-regular languages. In [12], it has been shown that a language which is both  $\omega$ B- and  $\omega$ S-regular is also  $\omega$ -regular. We aim at providing a characterization of languages which are both  $\omega$ B- (resp.,  $\omega$ S-) and  $\omega$ T-regular. We are also interested in (modal) temporal logic counterparts of extended  $\omega$ -regular languages. To the best of our knowledge, none was provided in the literature. We started to fill such a gap in [14, 15].

**Acknowledgements.** We would like to acknowledge prof. Massimo Benerecetti for some helpful comments on the relationship between  $*-$ ,  $B-$ ,  $S-$ , and  $T$ -constructors.

## References

- [1] R. Alur & T. A. Henzinger (1998): *Finitary Fairness*. *ACM Trans. Program. Lang. Syst.* 20(6), pp. 1171–1194, doi:10.1145/295656.295659.
- [2] D. Barozzini, D. Della Monica, A. Montanari & P. Sala: *Extending  $\omega$ -regular languages with a strong  $T$ -constructor:  $\omega T$ -regular languages and counter-queue automata*. Research Report 2017/01, Dept. of Mathematics, Computer Science, and Physics, University of Udine, Italy.
- [3] M. Bojańczyk (2004): *A bounding quantifier*. In: *CSL, LNCS 3210*, Springer, pp. 41–55, doi:10.1007/978-3-540-30124-0\_7.
- [4] M. Bojańczyk (2011): *Weak MSO with the Unbounding Quantifier*. *Theory of Computing Systems* 48(3), pp. 554–576, doi:10.1007/s00224-010-9279-2.
- [5] M. Bojańczyk & T. Colcombet (2006): *Bounds in  $\omega$ -Regularity*. In: *LICS*, pp. 285–296, doi:10.1109/LICS.2006.17.
- [6] M. Bojańczyk, P. Parys & S. Toruńczyk (2016): *The MSO+U Theory of  $(N, <)$  Is Undecidable*. In: *STACS, LIPIcs 47*, pp. 21:1–21:8, doi:10.4230/LIPIcs.STACS.2016.21.
- [7] J. R. Büchi (1962): *On a decision method in restricted second order arithmetic*. In: *Proc. of the 1960 Int. Congress on Logic, Methodology and Philosophy of Science*, pp. 1–11.
- [8] D. Della Monica, A. Montanari, A. Murano & P. Sala (2016): *Prompt Interval Temporal Logic*. In: *JELIA, LNCS 10021*, Springer, pp. 207–222, doi:10.1007/978-3-319-48758-8\_14.
- [9] C. C. Elgot & M. O. Rabin (1966): *Decidability and Undecidability of Extensions of Second (First) Order Theory of (Generalized) Successor*. *J. Symb. Log.* 31(2), pp. 169–181, doi:10.1002/malq.19600060105.
- [10] S. Hummel & M. Skrzypczak (2012): *The Topological Complexity of MSO+U and Related Automata Models*. *Fundam. Inform.* 119(1), pp. 87–111, doi:10.3233/FI-2012-728.
- [11] O. Kupferman, N. Piterman & M. Y. Vardi (2009): *From liveness to promptness*. *Formal Methods in System Design* 34(2), pp. 83–103, doi:10.1007/s10703-009-0067-z.
- [12] M. Skrzypczak (2014): *Separation Property for  $\omega B$ - and  $\omega S$ -regular Languages*. *Logical Methods in Computer Science* 10(1), doi:10.2168/LMCS-10(1:8)2014.
- [13] R. McNaughton (1966): *Testing and Generating Infinite Sequences by a Finite Automaton*. *Information and Control* 9(5), pp. 521–530, doi:10.1016/S0019-9958(66)80013-X.
- [14] A. Montanari & P. Sala (2013): *Adding an equivalence relation to the interval logic  $AB\bar{B}$ : complexity and expressiveness*. In: *LICS*, IEEE Computer Society, pp. 193–202, doi:10.1109/LICS.2013.25.
- [15] A. Montanari & P. Sala (2013): *Interval Logics and  $\omega B$ -Regular Languages*. In: *LATA, LNCS 7810*, Springer, pp. 431–443. Available at [http://dx.doi.org/10.1007/978-3-642-37064-9\\_38](http://dx.doi.org/10.1007/978-3-642-37064-9_38).