

UNIVERSAL QUANTIFICATION MAKES AUTOMATIC STRUCTURES HARD TO DECIDE

CHRISTOPH HAASE ^a AND RADOSŁAW PIÓRKOWSKI ^{a,b}

^a Department of Computer Science, University of Oxford, Oxford, United Kingdom

^b Department of Computer Science, University of Warwick, Coventry, United Kingdom

ABSTRACT. Automatic structures are first-order structures whose universe and relations can be represented as regular languages. It follows from the standard closure properties of regular languages that the first-order theory of an automatic structure is decidable. While existential quantifiers can be eliminated in linear time by application of a homomorphism, universal quantifiers are commonly eliminated via the identity $\forall x. \Phi \equiv \neg(\exists x. \neg\Phi)$. If Φ is represented in the standard way as an NFA, a priori this approach results in a doubly exponential blow-up. However, the recent literature has shown that there are classes of automatic structures for which universal quantifiers can be eliminated by different means without this blow-up by treating them as first-class citizens and not resorting to double complementation. While existing lower bounds for some classes of automatic structures show that a singly exponential blow-up is unavoidable when eliminating a universal quantifier, it is not known whether there may be better approaches that avoid the naïve doubly exponential blow-up, perhaps at least in restricted settings.

In this paper, we answer this question negatively and show that there is a family of NFA representing automatic relations for which the minimal NFA recognising the language after eliminating a single universal quantifier is doubly exponential, and deciding whether this language is empty is EXPSpace-complete.

The techniques underlying our EXPSpace lower bound further enable us to establish new lower bounds for some fragments of Büchi arithmetic with a fixed number of quantifier alternations.

1. INTRODUCTION

Quantifier elimination is a standard technique to decide logical theories. A logical theory \mathcal{T} admits quantifier elimination whenever for every quantifier free conjunction of literals $\Phi(x, y_1, \dots, y_n)$ of \mathcal{T} there is a quantifier free formula $\Psi(y_1, \dots, y_n)$ such that $\mathcal{T} \models \exists x. \Phi \leftrightarrow \Psi$. Universal quantifiers can then be eliminated simply by applying the duality $\forall x. \Phi \equiv \neg(\exists x. \neg\Phi)$. If the formula Ψ above is effectively computable then \mathcal{T} is decidable provided that its quantifier-free fragment is decidable. For quantifier elimination procedures, the computationally most expensive step is the elimination of an existential quantifier, since negating a formula can be performed on a syntactic level.

Automatic structures [Hod82, KN95, BG00] are a family of first-order structures whose corresponding first-order theory can be decided using automata-theoretic methods, as an alternative approach to syntactic quantifier elimination. In their simplest variant, automatic

Key words and phrases: automatic structures, universal projection, quantifier elimination, Büchi arithmetic.

structures are relational first-order structures whose universe is isomorphic to a regular language $L \subseteq \Sigma^*$ over some alphabet Σ , and whose n -ary relations are interpreted as regular languages over $(\Sigma^n)^*$. It follows that the set of all satisfying assignments of a quantifier-free formula $\Phi(x_1, \dots, x_{m+1})$ can be obtained as the language $\mathcal{L}(\mathcal{A}) \subseteq (\Sigma^{m+1})^*$ of some finite-state automaton \mathcal{A} . In this setting, eliminating existential quantifiers is easy. In order to obtain a finite-state automaton whose language encodes the satisfying assignments to $\exists x_{m+1}. \Phi$, it suffices to apply the homomorphism induced by the mapping $h: (\Sigma^{m+1}) \rightarrow (\Sigma^m)$ such that $h(u_1, \dots, u_{m+1}) := (u_1, \dots, u_m)$ to $\mathcal{L}(\mathcal{A})$. This can be performed in linear time, even when \mathcal{A} is non-deterministic. However, if \mathcal{A} is non-deterministic then computing a finite-state automaton whose language encodes the complement of Φ is computationally difficult and may lead to an automaton with $2^{\Omega(|\mathcal{A}|)}$ states. In particular, due to double complementation, eliminating a universal quantifier may *a priori* lead to an automaton with $2^{2^{\Omega(|\mathcal{A}|)}}$ states. Notable examples of automatic structures are Presburger arithmetic [Pre29], the first-order theory of the structure $\langle \mathbb{N}, 0, 1, +, = \rangle$, and its extension Büchi arithmetic [Büc60, Bru85, BHMV94]. Tool suites such as LASH [Boi], TAPAS [LP09] and WALNUT [Mou16] are based on the automata-theoretic approach and have successfully been used to decide challenging instances of Presburger arithmetic and Büchi arithmetic from various application domains. Those tools eliminate universal quantifiers via double complementation.

Yet another approach to deciding Presburger arithmetic is based on manipulating semi-linear sets [GS66, CHM22], which are generalisations of ultimately periodic sets to arbitrary tuples of integers in \mathbb{N}^d . They are similar to automata-based methods in terms of the computational difficulty of existential projection and complementation: the former is easy whereas the latter is difficult.

For certain classes of automatic structures, it is possible to avoid eliminating universal quantifiers via existential projection and negation. For example, it was shown in [CH17] that deciding sentences of quantified integer programming $\exists \bar{x}_1 \forall \bar{x}_2 \dots \exists \bar{x}_n. A \cdot \bar{x} \geq \bar{b}$ is complete for the n th level of the polynomial hierarchy. The upper bound was obtained by manipulating so-called hybrid linear sets, which characterise the sets of integer solutions of systems of linear equations $A \cdot \bar{x} \geq \bar{b}$. A key technique introduced in [CH17] is called *universal projection* and enables directly eliminating universal quantifiers instead of resorting to double complementation and existential projection. Given $S \subseteq \mathbb{N}^{d+k}$, the universal projection of S onto the first d coordinates is defined as

$$\pi_d^{\forall}(S) := \{ \bar{u} \in \mathbb{N}^d \mid (\bar{u}, \bar{v}) \in S \text{ for all } \bar{v} \in \mathbb{N}^k \}.$$

It is shown in [CH17] that if S is a hybrid linear set then $\pi_d^{\forall}(S)$ is a hybrid linear set that can be obtained as a finite intersection of existential projections of certain hybrid linear sets. Moreover, the growth of the constants in the description of the hybrid linear set is only polynomial. Neither syntactic quantifier elimination nor automata-based methods are powerful enough to derive those tight upper bounds for quantified integer programming.

Another example is a recent paper of Boigelot et al. [BFV23] showing that, in an automata-theoretic approach for a fragment of Presburger arithmetic with uninterpreted predicates, a universal projection step can directly be carried out on the automata level without complementation and only results in a singly exponential blow-up.

Those positive algorithmic and structural results are specific to Presburger arithmetic and leave open the option that it may be possible to establish analogous results for general automatic structures. The starting point of this paper is the question of whether, given

a non-deterministic finite automaton \mathcal{A} whose language $\mathcal{L}(\mathcal{A}) \subseteq (\Sigma^{d+k})^*$ encodes the set of solutions of some quantifier-free formula Φ , there is a more efficient way to eliminate a (block of) universally quantified variable(s) than to first complement \mathcal{A} , next to perform an existential projection step, and finally to complement the resulting automaton again, especially in the light of the results of [CH17, CHM22]. Such a method would have direct consequences for tools such as WALNUT which perform the aforementioned sequence of operations in order to eliminate universal quantifiers. In particular, WALNUT is not restricted to automata resulting from formulas of linear arithmetic and allows users to directly specify a finite-state automaton when desired.

For better or worse, however, as the main result of this paper, we show that deciding whether the universal projection $\pi_d^{\forall}(\mathcal{L}(\mathcal{A}))$ of some regular language $\mathcal{L}(\mathcal{A}) \subseteq (\Sigma^{d+k})^*$ is empty is complete for EXPSPACE. In particular, the lower bound already holds for $d = k = 1$, meaning that, in general, even for fixed-variable fragments of automatic structures, there is no algorithmically more efficient way to eliminate a single universal quantifier than the naïve one. The challenging part is to show the EXPSPACE lower bound, which requires an involved reduction from a tiling problem. This reduction also enables us to show that there is a family $(\mathcal{A}_n)_{n \in \mathbb{N}}$ of non-deterministic finite automata such that $|\mathcal{A}_n| = O(n^4)$ and the smallest non-deterministic finite automaton recognising the universal projection of $\mathcal{L}(\mathcal{A}_n)$ has $\Omega(2^{2^n})$ states.

One of the most prominent automatic structures is Büchi arithmetic [Büc60, Bru85]. Given an integer $p \geq 2$, *Büchi arithmetic of base p* is the first-order theory of the structure $\langle \mathbb{N}, 0, 1, +, V_p \rangle$, where V_p is a binary predicate such that $V_p(x, y)$ holds whenever x is the largest power of p dividing y without remainder, i.e., $x = p^k$ for some $k \geq 0$, $x \mid y$ and $px \nmid y$. Büchi arithmetic is a universal automatic structure in the following sense: For any regular language $L \subseteq (\Sigma^n)^*$, there is a formula $\Phi(\mathbf{x})$ of Büchi arithmetic with an $\exists^* \forall^*$ quantifier prefix such that the set of satisfying $\mathbf{x} \in \mathbb{N}^n$ is isomorphic to L [HR21]. The existential fragment of Büchi arithmetic is NP-complete, and the full first-order theory is complete for TOWER [GHW19] (see e.g. [Sch16] for a definition of TOWER). While the computational complexity of Presburger arithmetic with fixed quantifier alternation prefixes is well understood [Haa14], to the best of the authors' knowledge no (stronger) lower bounds are known when generalizing to Büchi arithmetic. A further contribution of this paper is to show that Büchi arithmetic with an $\exists^* \forall^* \exists^*$ quantifier prefix is EXPSPACE-hard, and it is 2-EXPSPACE-hard with a $\exists^* \forall^* \exists^* \forall^*$ quantifier prefix. Those lower bounds are obtained by adapting the aforementioned EXPSPACE lower bound for universal projection emptiness.

2. PRELIMINARIES

2.1. Regular languages and their compositions. For a word $w = a_1 a_2 \cdots a_n \in \Sigma^*$, we write $w[i]$ to denote its i th letter a_i , and $w[i, j]$ to denote the infix $a_i a_{i+1} \cdots a_j$ ($i \leq j$). We write $|w|$ for the length of w . A *proper suffix* of w is any infix $w[i, n]$ for some $1 < i \leq n$.

Regular expressions. A *regular expression* over the alphabet Σ is a term featuring Kleene star, concatenation and union operations, as well as \emptyset and all symbols from Σ as constants:

$$\mathcal{E}, \mathcal{E}' ::= \mathcal{E}^* \mid \mathcal{E} \cdot \mathcal{E}' \mid \mathcal{E} + \mathcal{E}' \mid \emptyset \mid a \text{ for every } a \in \Sigma$$

For notational convenience, we also use sets of symbols $A \subseteq \Sigma$ as constants, and a k -fold concatenation \mathcal{E}^k for every $k \in \mathbb{N}$; we also drop the concatenation dot most of the time. The

language $\mathcal{L}(\mathcal{E}) \subseteq \Sigma^*$ is defined by structural induction, by defining $\mathcal{L}(\emptyset) := \emptyset$ and $\mathcal{L}(a) := \{a\}$, and using the standard semantics of the three operations. The class of languages definable by regular expressions is called *regular languages*. The size $|\mathcal{E}|$ of a regular expression \mathcal{E} is defined recursively as 1 plus the sizes of its subexpressions, where $|a| = |\emptyset| := 1$. For $\rho: \Sigma \rightarrow \Gamma$ and a regular expression \mathcal{E} , $\rho(\mathcal{E})$ is a regular expression over Γ obtained through substituting every constant $a \in \Sigma$ appearing in \mathcal{E} by $\rho(a)$.

Finite-state automata. Regular languages can also be represented by *non-deterministic finite-state automata* (NFA). Such an automaton is a tuple $\mathcal{A} = (Q, \Sigma, \delta, Q_I, Q_F)$, where Q is a finite non-empty set of *states*, Σ is a finite *alphabet*, $\delta \subseteq Q \times \Sigma \times Q$ is the *transition relation*, $Q_I \subseteq Q$ is the set of *initial states*, and $Q_F \subseteq Q$ is the set of *final states*. A triple $(p, a, q) \in Q \times \Sigma \times Q$ is called a *transition* and denoted as $p \xrightarrow{a} q$. A *run* of \mathcal{A} from a state q_0 to a state q_n ($n \in \mathbb{N}$) on a word $w = a_1 a_2 \cdots a_n \in \Sigma^*$ is a finite sequence of transitions $(q_{i-1} \xrightarrow{a_i} q_i)_{1 \leq i \leq n}$ such that $q_{i-1} \xrightarrow{a_i} q_i \in \delta$ for every i . A word $w \in \Sigma^*$ is *accepted* by \mathcal{A} if there exists a run of \mathcal{A} from some $q_I \in Q_I$ to $q_F \in Q_F$ over w . The *language* of \mathcal{A} is defined as $\mathcal{L}(\mathcal{A}) := \{w \in \Sigma^* \mid w \text{ is accepted by } \mathcal{A}\}$. We define the size of \mathcal{A} as $|\mathcal{A}| := |Q| + |\Sigma| + |\delta|$. Subsequently, we will implicitly apply the well-known fact that the size of an NFA accepting the complement of $\mathcal{L}(\mathcal{A})$ is $O(2^{|\mathcal{A}|} \cdot |\Sigma|)$, and that it has $2^{|\mathcal{A}|}$ states.

Below we state, without proofs, a few folklore properties of NFA:

Fact 2.1 (NFA closed under language union). For any NFA \mathcal{A}, \mathcal{B} over Γ with states $Q_{\mathcal{A}}, Q_{\mathcal{B}}$, there exists an NFA $\mathcal{A} \oplus \mathcal{B}$ with $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$ states such that $\mathcal{L}(\mathcal{A} \oplus \mathcal{B}) = \mathcal{L}(\mathcal{A}) \cup \mathcal{L}(\mathcal{B})$.

Given finite alphabets Σ, Γ , a *homomorphism* is a function $\rho: \Sigma^* \rightarrow \Gamma^*$ such that $\rho(v \cdot w) = \rho(v) \cdot \rho(w)$ for all $v, w \in \Sigma^*$. It follows that ρ is fully defined by specifying $\rho(a)$ for all $a \in \Sigma$.

Example 2.2. Let $\Sigma = \{a, b\}$, $\Gamma = \{x, y\}$, and define ρ such that $\rho(a) = xy$ and $\rho(b) = \epsilon$. Then $\rho(abba) = xyxy$ and $\rho^{-1}(\{xyxy\}) = \mathcal{L}(b^*ab^*ab^*)$.

Fact 2.3 (NFA closed under inverse homomorphisms). For any NFA $\mathcal{A} = (Q, \Sigma, \delta, Q_I, Q_F)$ and a homomorphism $\rho: \Sigma^* \rightarrow \Gamma^*$, there exists an NFA $\rho^{-1}(\mathcal{A})$ with $|Q|$ states such that $\mathcal{L}(\rho^{-1}(\mathcal{A})) = \rho^{-1}(\mathcal{L}(\mathcal{A}))$.

Fact 2.4 (NFA closed under concatenation of languages). For any NFA \mathcal{A}, \mathcal{B} with states $Q_{\mathcal{A}}$ and $Q_{\mathcal{B}}$, respectively, there exists an NFA $\mathcal{A} \odot \mathcal{B}$ with $|Q_{\mathcal{A}}| + |Q_{\mathcal{B}}|$ states such that $\mathcal{L}(\mathcal{A} \odot \mathcal{B}) = \mathcal{L}(\mathcal{A}) \cdot \mathcal{L}(\mathcal{B}) := \{u \cdot v \mid u \in \mathcal{L}(\mathcal{A}) \text{ and } v \in \mathcal{L}(\mathcal{B})\}$.

Fact 2.5 (translating regular expressions to NFA). There is a deterministic algorithm [Tho68] that, given a regular expression \mathcal{E} , constructs an NFA $\mathcal{A}(\mathcal{E})$ such that $\mathcal{L}(\mathcal{A}(\mathcal{E})) = \mathcal{L}(\mathcal{E})$.

Filters. A *filter* is an auxiliary term introduced to simplify the proofs in Section 3, allowing for a modular design of regular languages. Fix a finite alphabet Σ and let $\Phi := \{\top, \perp\}$. Define homomorphisms $\psi_{\text{in}}, \psi_{\text{out}}: (\Sigma \times \Phi)^* \rightarrow \Sigma^*$ by their actions on a single letter

$$\begin{aligned} \psi_{\text{in}}(a, b) &:= a & \psi_{\text{out}}(a, \top) &:= a & \psi_{\text{out}}(a, \perp) &:= \epsilon. \\ \text{(output every symbol from } \Sigma) & & \text{(output only symbols paired with } \top) & & & \end{aligned}$$

A filter over an alphabet Σ is any language $F \subseteq (\Sigma \times \Phi)^*$. It induces a binary *input-output relation* $\mathcal{R}(F) \subseteq \Sigma^* \times \Sigma^*$ between input words u and their subsequences v :

$$(u, v) \in \mathcal{R}(F) \iff u = \psi_{\text{in}}(w) \text{ and } v = \psi_{\text{out}}(w) \text{ for some } w \in F.$$

We define $F(u) := \{v \mid (u, v) \in \mathcal{R}(F)\}$ to be the set of all possible outputs of F on u .

Filtering regular expressions. A *filtering regular expression* \mathcal{F} over alphabet Σ is any regular expression over $\Sigma \times \Phi$. We write $\mathcal{F}(w) := \mathcal{L}(\mathcal{F})(w)$. To simplify the notation, we only write the Σ component of the constants, and underline parts of the expression. A symbol a appearing in an underlined fragment represents a pair (a, \top) , and in a fragment which is not underlined a pair (a, \perp) . Intuitively, underlined portions correspond to parts of the words being output. We apply the same notational convention to words $w \in (\Sigma \times \Phi)^*$. Additionally, for $\rho: \Sigma \rightarrow \Gamma$, we abuse the notation and extend it to the naturally defined homomorphism of type $\Sigma \times \Phi \rightarrow \Gamma \times \Phi$, which just preserves the coordinate belonging to Φ .

Example 2.6. Fix $A = \{a, b, c, \dots, z\}$. Consider a filtering regular expression \mathcal{F} and a word w , both over $A \cup \{\perp\}$:

$$\mathcal{F} := (\underline{A} A^* \perp)^* \underline{A} A^* \quad w := \text{nondeterministic_finite_automaton.}$$

We have:

$$\begin{aligned} \mathcal{F}(w) &= \{\text{nfa}\}, \\ \mathcal{F} &= ((A \times \{\top\}) \cdot (A \times \{\perp\})^* \cdot (\perp, \perp))^* \cdot (A \times \{\top\}) \cdot (A \times \{\perp\})^*, \\ \mathcal{L}(\mathcal{F}) &\ni \underline{\text{nondeterministic_finite_automaton}}. \end{aligned}$$

2.2. Automatic relations. Let Σ be a finite alphabet such that $\# \notin \Sigma$. We denote by $\Sigma_{\#} := \Sigma \cup \{\#\}$. Let $w_1, \dots, w_k \in \Sigma^*$ such that $w_i = a_{i,1}a_{i,2} \cdots a_{i,\ell_i}$, and $\ell := \max\{\ell_1, \dots, \ell_k\}$. For all $1 \leq i \leq k$ and $\ell_i < j \leq \ell$, set $a_{i,j} := \#$. The *convolution* $w_1 \otimes w_2 \otimes \cdots \otimes w_k$ of w_1, \dots, w_k is defined as

$$w_1 \otimes w_2 \otimes \cdots \otimes w_k := \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{k,1} \end{bmatrix} \begin{bmatrix} a_{1,2} \\ \vdots \\ a_{k,2} \end{bmatrix} \cdots \begin{bmatrix} a_{1,\ell} \\ \vdots \\ a_{k,\ell} \end{bmatrix} \subseteq (\Sigma_{\#}^k)^*.$$

For $R \subseteq (\Sigma^*)^k$ and $L \subseteq (\Sigma_{\#}^k)^*$ define

$$\begin{aligned} \text{Rel2Lang}(R) &:= \{w_1 \otimes w_2 \otimes \cdots \otimes w_k \mid (w_1, w_2, \dots, w_k) \in R\}, \\ \text{Lang2Rel}(L) &:= \{(w_1, w_2, \dots, w_k) \mid w_1 \otimes w_2 \otimes \cdots \otimes w_k \in L\}. \end{aligned}$$

In this paper, we say that a relation $R \subseteq (\Sigma^*)^k$ is *automatic* whenever $\text{Rel2Lang}(R)$ is regular. Furthermore, we assume that $\text{Rel2Lang}(R)$ is given by some NFA $\mathcal{A}_R = (Q, \Sigma_{\#}^k, \delta, Q_I, Q_F)$.

Clearly, not every NFA $\mathcal{A} = (Q, \Sigma_{\#}^k, \delta, Q_I, Q_F)$ is associated with an automatic relation $R \subseteq \Sigma^k$ since there are *a priori* no restrictions on the occurrences of the padding symbol “#”. The language $L_{\times} \subseteq (\Sigma_{\#}^k)^*$ of all incorrect words that cannot be obtained as a convolution of words $w_1, \dots, w_k \in \Sigma^*$ can be characterized by the following regular expression:

$$(\Sigma_{\#}^k)^* \cdot \left[\{\#\}^k + \sum_{1 \leq i \leq k} \left((\Sigma_{\#}^{i-1} \times \{\#\} \times \Sigma_{\#}^{k-i}) \cdot (\Sigma_{\#}^{i-1} \times \Sigma \times \Sigma_{\#}^{k-i}) \right) \right] \cdot (\Sigma_{\#}^k)^*.$$

This regular expression “guesses” that either a letter consisting solely of k # symbols occurs, or in some row of a word in $(\Sigma_{\#}^k)^*$ a “#” symbol is followed by a symbol in Σ . The language of this regular expression can be implemented by an NFA with $k + 2$ states. Hence, the complement $L_{\checkmark} := \overline{L_{\times}}$ of L_{\times} , characterizing all “good” words, can be recognized by an NFA

with 2^{k+2} states. For the sake of readability, we do not parametrise L_{\times} explicitly with k ; the relevant k will always be clear from the context.

The *existential projection* of $R \subseteq (\Sigma^*)^{d+k}$ onto the first d components is defined as

$$\pi_d^{\exists}(R) := \{\bar{u} \in (\Sigma^*)^d \mid (\bar{u}, \bar{w}) \in R \text{ for some } \bar{w} \in (\Sigma^*)^k\}.$$

The dual of existential projection is *universal projection*:

$$\pi_d^{\forall}(R) := \{\bar{u} \in (\Sigma^*)^d \mid (\bar{u}, \bar{w}) \in R \text{ for all } \bar{w} \in (\Sigma^*)^k\}.$$

It is clear that $\pi_d^{\forall}(R) = \overline{\pi_d^{\exists}(\overline{R})}$. We overload the projection notation for languages

$$\pi_d^{\exists}(L) := \text{Rel2Lang}(\pi_d^{\exists}(\text{Lang2Rel}(L))) \quad \pi_d^{\forall}(L) := \text{Rel2Lang}(\pi_d^{\forall}(\text{Lang2Rel}(L))).$$

In this article, given \mathcal{A}_R such that $\text{Rel2Lang}(R) = \mathcal{L}(\mathcal{A}_R) \subseteq (\Sigma_{\#}^{d+k})^*$, we are concerned with the computational complexity of deciding whether $\pi_d^{\forall}(R) = \emptyset$, measured in terms of $|\mathcal{A}_R|$. In Sections 3 and 5 we will prove the following.

Theorem 2.7. *Deciding whether $\pi_d^{\forall}(R) \neq \emptyset$ for an automatic relation $R \subseteq (\Sigma^*)^{d+k}$ given by an NFA \mathcal{A}_R is EXPSPACE-complete. The lower bound already holds for $d = k = 1$.*

2.3. Büchi arithmetic. As discussed in the introduction, given an integer $p \geq 2$, *Büchi arithmetic of base p* is the first-order theory of the structure $\langle \mathbb{N}, 0, 1, +, V_p \rangle$, where V_p is a binary predicate such that $V_p(x, y)$ holds whenever x is the largest power of p dividing y without remainder, i.e., $x = p^k$ for some $k \geq 0$, $x \mid y$ and $px \nmid y$. Atomic formulas of Büchi arithmetic are V_p predicates or linear inequalities of the form $a_1 \cdot x_1 + \dots + a_n \cdot x_n \geq b$, where all numbers are encoded in binary. We use some standard abbreviations such as strict inequalities, etc. Note that $V_p(x, x)$ holds if and only if x is a power of p , and we write $P_p(x)$ to abbreviate $V_p(x, x)$. Given a formula $\Phi(x_1, \dots, x_n)$ of Büchi arithmetic, we define $\llbracket \Phi \rrbracket := \{(m_1, \dots, m_n) \in \mathbb{N}^n : \Phi(m_1, \dots, m_n) \text{ is valid}\}$ to be its set of satisfying assignments. Given a string $w = d_0 \dots d_n \in \{0, \dots, p-1\}^+$, we define $\llbracket w \rrbracket_p \in \mathbb{N} := \sum_{i=0}^n p^i \cdot d_i$.

3. EMPTINESS AFTER UNIVERSAL PROJECTION IS EXPSPACE-HARD

3.1. Tiling problems. Let $\mathcal{T} \subseteq_{\text{fin}} \mathbb{N}^4$ be a set of *tiles* with colours coded as four-tuples of numbers with associated projections *top*, *right*, *bottom*, *left*: $\mathbb{N}^4 \rightarrow \mathbb{N}$ to access individual colours of a tile, and let $\text{colours}(\mathcal{T}) := \text{top}(\mathcal{T}) \cup \text{right}(\mathcal{T}) \cup \text{bottom}(\mathcal{T}) \cup \text{left}(\mathcal{T})$.

Example 3.1. A tile $t = (2, 4, 3, 3)$ is depicted as  with various auxiliary background shades corresponding to colour values.

A \mathcal{T} -tiling of size $(h, w) \in \mathbb{N}_+^2$ is an $h \times w$ matrix $T = [t_{i,j}]_{i,j} \in \mathcal{T}^{h \times w}$. It is *valid* whenever colours of the neighbouring tiles match:

$$\text{bottom}(t_{i,j}) = \text{top}(t_{i+1,j}) \quad \text{for every } 1 \leq i \leq h-1 \text{ and } 1 \leq j \leq w, \quad (3.1)$$

$$\text{right}(t_{i,j}) = \text{left}(t_{i,j+1}) \quad \text{for every } 1 \leq i \leq h \quad \text{and } 1 \leq j \leq w-1. \quad (3.2)$$

See Figure 2 on Page 17 for an example of a valid tiling. A \mathcal{T} -tiling of width $w \in \mathbb{N}_+$ is any tiling in $\mathcal{T}^{h \times w}$ for some $h \in \mathbb{N}_+$. We define

$$\mathcal{T}^{\star \times w} := \bigcup_{h \in \mathbb{N}_+} \mathcal{T}^{h \times w}.$$

Additionally, for two distinguished tiles $t^\wedge, t^\vee \in \mathcal{T}$, let $(\mathcal{T}, t^\wedge, t^\vee)$ -tiling be any \mathcal{T} -tiling with t^\wedge placed in its top-left corner, and t^\vee in its bottom-right corner.

Problem 3.2. *CORRIDORTILING*

INPUT: A 4-tuple $(\mathcal{T}, t^\wedge, t^\vee, n)$, where

- $\mathcal{T} \subseteq_{\text{fin}} \mathbb{N}^4$ is a finite set of tiles,
- $t^\wedge, t^\vee \in \mathcal{T}$,
- $n \in \mathbb{N}_+$ given in unary.

QUESTION: Does there exist a valid $(\mathcal{T}, t^\wedge, t^\vee)$ -tiling of width 2^n ?

By $\mathbb{T} \subset \mathcal{P}_{\text{fin}}(\mathbb{N}^4) \times \mathbb{N}^4 \times \mathbb{N}^4 \times \mathbb{N}_+$ we denote the set of all syntactically valid instances of the above problem.

Lemma 3.3. *CORRIDORTILING is EXPSPACE-hard.*

Proof (Sketch). It is part of the folklore of the theory of computation that tiling problems can simulate the computation of Turing machines, the width of the requested tiling corresponding to the length of tape the machine is allowed to use. EXPSPACE-completeness of a variant similar to the one above is sketched in [Sch19], see also the CORRIDORTILING problem in [vEB19] and the discussion of it capturing computations of space-bounded Turing machines.

In more technical terms, let M be a Turing machine using at most $2^{p(m)}$ tape cells on an input $w \in \{0, 1\}^*$ of length m for some polynomial $p: \mathbb{N} \rightarrow \mathbb{N}$. To decide whether M accepts w , we show how to construct in logarithmic space an instance $(\mathcal{T}, t^\wedge, t^\vee, n)$ of *CORRIDORTILING* from an encoding of M and w such that a valid tiling exists if and only if M has an accepting computation on input w . Note that any computation of M can be represented by a computation table of width $2^{p(m)}$ such that the i th row of that table represents the full configuration of M after i computation steps. Each cell of the computation table is either a symbol of the working tape, or such a symbol paired with a control state q of M to indicate that M is currently reading that symbol in state q . Consistency of a computation table can be verified locally: in every row, only one cell paired with some q is allowed to occur, and between rows, the movement of the head, the state transition and the updating of cells occur only locally around the head position. This local consistency property is precisely what allows the computation table to be enforced by local tiling constraints in a *CORRIDORTILING* instance. Finally, requiring the initial state of M to appear in the first row and the accepting state in the last row of the computation table ensures that the table represents an accepting computation of M .

To define $(\mathcal{T}, t^\wedge, t^\vee, n)$, we set $n = p(m) + 1$ and choose a finite set of tiles that allow us to represent a computation table as described above. The horizontal adjacency constraints enable us to adhere to the requirement that a state appears only once along every row, and vertical constraints can be used to ensure correct state, head, and tape updates between consecutive computation steps. The special tile t^\wedge is chosen such that the first row of a tiling represents a valid initial configuration of M reading w , and whenever an accepting state occurs in some row, the tiles are chosen such that t^\vee can appear at the very last tile of the tiling. □

3.2. The reduction. We prove Theorem 2.7 by a reduction from *CORRIDORTILING* and show that EXPSPACE-hardness occurs in the simplest case of universal projection: projecting a binary relation to get a unary one. Intuitively, for each instance $\mathcal{I} = (\mathcal{T}, t^\wedge, t^\vee, n)$ of *CORRIDORTILING*, we construct an automaton $\mathcal{A}_{\mathcal{I}}$ such that $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{I}}))$ is not empty if and only if \mathcal{I} is a YES-instance. Formally, we provide a family of LOGSPACE-constructible NFA $(\mathcal{A}_{\mathcal{I}})_{\mathcal{I} \in \mathbb{T}}$ over the alphabet $(\Sigma_{\mathcal{I}} \cup \{\#\})^2$ for some $\Sigma_{\mathcal{I}}$, each of size $O(n^4 \cdot |\mathcal{T}|^3)$, representing the relation $Lang2Rel(\mathcal{L}(\mathcal{A}_{\mathcal{I}})) \subseteq (\Sigma_{\mathcal{I}}^*)^2$, such that

$$\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{I}})) \neq \emptyset \iff \text{there exists a valid } (\mathcal{T}, t^\wedge, t^\vee)\text{-tiling of width } 2^n. \quad (3.3)$$

The LOGSPACE constructibility is easy to verify and we chose not to provide explicit arguments for it. For the remainder of this section, we fix an instance $\mathcal{I} \in \mathbb{T}$. For technical reasons, we assume with no loss of generality that $n \geq 6$.

In Section 3.3, we define $\Sigma_{\mathcal{I}}$, specify a language $L_{\mathcal{I}} \in \Sigma_{\mathcal{I}}^*$, and prove that:

Lemma 3.4. $L_{\mathcal{I}} \neq \emptyset \iff$ *there exists a valid $(\mathcal{T}, t^\wedge, t^\vee)$ -tiling of width 2^n .*

In turn in Section 3.4, we construct an NFA $\mathcal{A}_{\mathcal{I}}$ such that

Lemma 3.5. $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{I}})) = L_{\mathcal{I}}$.

This completes the proof of Theorem 2.7, the correctness of the reduction stemming directly from Lemmas 3.4 and 3.5.

3.3. Word encoding of tilings. Here, we provide $\Sigma_{\mathcal{I}}$ and an encoding $enc_{\mathcal{I}}: \mathcal{T}^{\star \times 2^n} \rightarrow \Sigma_{\mathcal{I}}^*$. Then we define $L_{\mathcal{I}}$ as an intersection of six languages, and prove Lemma 3.4 by showing that it coincides with the language of encodings of valid tilings.

Let $N_n := \mathbb{N} \cap [0, n]$. Additionally, let $N_n^{\sim k} := \{i \in N_n \mid i \sim k\}$ for $\sim \in \{<, =, >\}$ and $k \in \mathbb{N}$. The alphabet $\Sigma_{\mathcal{I}}$ consists of three groups of symbols – tiles from \mathcal{T} , numbers from N_n , and auxiliary symbols:

$$\Sigma_{\mathcal{I}} := \mathcal{T} \cup N_n \cup \{\mathbf{A}, \mathbb{I}, \mathbb{I}, \langle, \rangle\}.$$

The symbol **A** is a mnemonic marking places where we enforce “for-all”-type properties. In what follows, we colour some symbols (e.g., **3010** t **20103**) to assist in understanding the construction; such designations are auxiliary and are not reflected in the alphabet. The encoding of runs uses the word $COMB_n \in N_n^*$

$$COMB_n := n COMB'_{n-1} n,$$

where the words $(COMB'_i)_{0 \leq i \leq n}$ are defined recursively as

$$COMB'_0 := 0$$

$$COMB'_i := COMB'_{i-1} i COMB'_{i-1} \quad \text{for } 0 < i \leq n.$$

Observe that $COMB_n$ has length exactly $2^n + 1$. The recursive definition of $COMB_n$ will later enable us to recognise $\{COMB_n\}$ as an intersection of n NFA, each with a constant number of states. The intersection itself will be implemented as a universal projection step.

Example 3.6. $COMB_4$ is 40102010301020104 and has length 17.

The following observation shows that $COMB_n$ is tightly related to a binary counter on 2^n bits, which will play a crucial role in Section 6.2.

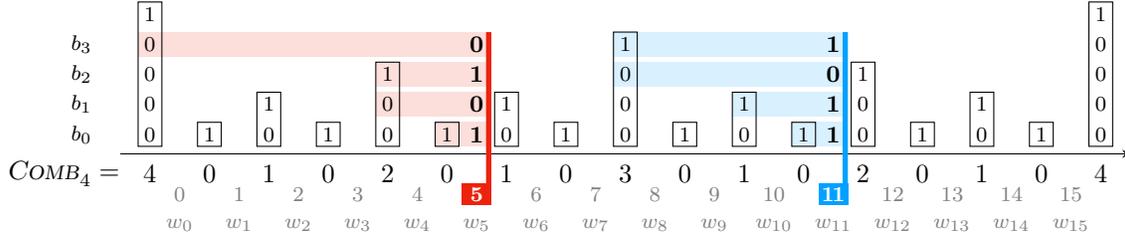


FIGURE 1. Illustration of $COMB_4$ and how to decode positions in a comb.

Observation 3.7. Consider a 2^n -bit counter counting from 0 to $2^{2^n} - 1$ and overflowing back to 0:

$$\underbrace{0 \dots 00}_{2^n \text{ times}} \xrightarrow{0} 0 \dots 01 \xrightarrow{1} 0 \dots 10 \xrightarrow{0} 0 \dots 11 \xrightarrow{2} \dots \rightarrow 1 \dots 10 \xrightarrow{1} \underbrace{1 \dots 11}_{2^n \text{ times}} \xrightarrow{N} \underbrace{0 \dots 00}_{2^n \text{ times}}$$

Let w_i be the binary representation of i on 2^n bits; set $w_{2^{2^n}} := w_0$. We index bits of w_i starting from the least significant one (index 0) to the most significant one (index $2^n - 1$). The i th letter of $COMB_{2^n}$ corresponds to the smallest index at which w_i and w_{i+1} differ for $i \in \{0, \dots, 2^{2^n} - 1\}$. Furthermore, for every such i , the j th bit of w_i is 1 if and only if there is a position i' to the left of i such that $COMB_{2^n}[i'] = j$ and the infix $COMB_{2^n}[i' + 1, i]$ contains only symbols smaller than j .

An example involving $COMB_4 = 40102 \dots 04$ is given in Fig. 1. We index the positions between the letters of $COMB_4$, starting from 0 and displayed in a smaller grey font. Those positions correspond to the binary strings w_i above. We highlight positions 5 and 11 and illustrate how to obtain the binary representations w_5 and w_{11} based on the properties of the comb. To this end, above every element k of the comb, we vertically (top to bottom) write the binary string $1 \cdot 0^k$. In order to obtain, e.g. bit b_1 of the bit string $w_5 = b_3b_2b_1b_0$, we search to the left for the closest symbol “visible” at level 1, which is a 0 from the bit string 100, and which determines the value of $b_1 := 0$. Likewise, the most significant bit of w_{11} is 1 since at position 8, $COMB_4$ takes value 3 and is hence labelled with 1000.

We define the *encoding* function $enc_{\mathcal{T}}: \mathcal{T}^{* \times 2^n} \rightarrow \Sigma_{\mathcal{T}}^*$ in three steps. Let $T = [t_{i,j}]_{i,j} \in \mathcal{T}^{h \times 2^n}$ for some $h \in \mathbb{N}$. The tile $t_{i,j}$ in T is represented as

$$encCell_{\mathcal{T}}(T, i, j) := \langle COMB_n[1, j] t_{i,j} COMB_n[j + 1, 2^n + 1] \mathbf{A} \rangle,$$

a single row is encoded as

$$encRow_{\mathcal{T}}(T, i) := \mathbb{I} \prod_{1 \leq j \leq 2^n} encCell_{\mathcal{T}}(T, i, j) \mathbb{I},$$

and finally, the encoding of the entire tiling is defined as

$$enc_{\mathcal{T}}(T) := \mathbf{A} \prod_{1 \leq i \leq h} encRow_{\mathcal{T}}(T, i).$$

Example 3.8. The tiling $T = [t_{i,j}]_{i,j}$ of size $(2, 2^4)$ is encoded as

$$\mathbf{A} \langle \langle 4 t_{1,1} 0102010301020104 \mathbf{A} \rangle \dots \langle 40102 t_{1,5} 01030 \dots 04 \mathbf{A} \rangle \dots \langle 4010201030102010 t_{1,16} 4 \mathbf{A} \rangle \rangle \cdot \\ \cdot \mathbb{I} \langle \langle 4 t_{2,1} 0102010301020104 \mathbf{A} \rangle \dots \langle 40102 t_{2,5} 01030 \dots 04 \mathbf{A} \rangle \dots \langle 4010201030102010 t_{2,16} 4 \mathbf{A} \rangle \rangle \mathbb{I}.$$

The word above is written in two lines to make the correspondence to a tiling more apparent.

Languages of encodings. Define the language of encodings of valid tilings of width 2^n with t^\wedge, t^\vee in the correct corners

$$\text{VALIDENC}_{\mathcal{T}} := \{ \text{enc}_{\mathcal{T}}(T) \mid T \text{ is a valid } (\mathcal{T}, t^\wedge, t^\vee)\text{-tiling of width } 2^n \}.$$

In order to express the notion of an encoding of a valid tiling in a more tangible way, below we define languages $\text{COND}_{\mathcal{T}}^1, \dots, \text{COND}_{\mathcal{T}}^6$, which, as we prove in Lemma 3.9, jointly characterise encodings. The first three are given by regular expressions fixing basic properties of the encoding, the next two guarantee an appropriate width of the encoding, while the last one enforces in a non-trivial way that the colours match vertically.

Condition 1. Recall that the alphabet $\Sigma_{\mathcal{T}}$ is parametrised by n . Language $\text{COND}_{\mathcal{T}}^1$ is given by the regular expression

$$\mathcal{E}_{\mathcal{T}}^1 := \left(\llbracket \langle n \mathcal{T} N_n^* \mathbf{A} \rangle \left(\langle N_n^* \mathcal{T} N_n^* \mathbf{A} \rangle \right)^* \langle N_n^* \mathcal{T} n \mathbf{A} \rangle \rrbracket \right)^*.$$

Intuitively, encodings consist of rows bounded by \llbracket and \rrbracket ; each row comprises cells delimited by \langle and \rangle . The first cell begins with the number n followed by a tile, while the last one ends with a tile, n and \mathbf{A} . Observe that while the size of $\mathcal{B}_{\mathcal{T}}^1 := \mathcal{A}(\mathcal{E}_{\mathcal{T}}^1)$ grows linearly in n , its number of states is constant.

Condition 2. The language $\text{COND}_{\mathcal{T}}^2$ is defined by the regular expression

$$\mathcal{E}_{\mathcal{T}}^2 := \llbracket \langle N_n^* t^\wedge N_n^* \mathbf{A} \rangle \left(\langle N_n^* \mathcal{T} N_n^* \mathbf{A} \rangle \right)^* \rrbracket \Sigma_{\mathcal{T}}^* \llbracket \left(\langle N_n^* \mathcal{T} N_n^* \mathbf{A} \rangle \right)^* \langle N_n^* t^\vee N_n^* \mathbf{A} \rangle \rrbracket.$$

This requires the first row of a purported tiling to begin with t^\wedge , and the last row to end with t^\vee . As in Condition 1, $\text{COND}_{\mathcal{T}}^2$ is obtained by an NFA $\mathcal{B}_{\mathcal{T}}^2 := \mathcal{A}(\mathcal{E}_{\mathcal{T}}^2)$ with $O(1)$ states.

Condition 3. Let $Q := \text{colours}(\mathcal{T})$ and $\mathcal{B}_{\mathcal{T}}^3 := (Q, \Sigma_{\mathcal{T}}, \delta, Q, Q)$, where δ has transitions

$$\begin{aligned} i &\xrightarrow{t} j && \text{for every } i, j \in Q \text{ and } t \in \mathcal{T} \text{ such that } \text{left}(t) = i \text{ and } \text{right}(t) = j, \\ i &\xrightarrow{a} i && \text{for every } i \in Q \text{ and } a \in \Sigma_{\mathcal{T}} \setminus (\mathcal{T} \cup \{\llbracket\rrbracket\}), \\ i &\xrightarrow{\llbracket} j && \text{for every } i, j \in Q. \end{aligned}$$

We set $\text{COND}_{\mathcal{T}}^3 := \mathcal{L}(\mathcal{B}_{\mathcal{T}}^3)$; it contains encodings where tile colours match horizontally. Observe that $|Q| = O(|\mathcal{T}|)$.

Condition 4 (each cell contains a COMB_n). The definition of $\text{COND}_{\mathcal{T}}^4$ uses a filtering regular expression $\mathcal{F}_{\mathcal{T}}^4$:

$$\mathcal{F}_{\mathcal{T}}^4 := \langle \underline{N_n^*} \mathcal{T} \underline{N_n^*} \mathbf{A} \rangle \Sigma_{\mathcal{T}}^*$$

$$\text{COND}_{\mathcal{T}}^4 := \{ w \in \Sigma_{\mathcal{T}}^* \mid \text{COMB}_n \mathbf{A} \in \mathcal{F}_{\mathcal{T}}^4(v) \text{ for every proper suffix } v \text{ of } w \text{ such that } v[1] = \langle \}$$

Condition 5 (prefix of a cell and first symbols of following cells' suffixes form a COMB_n).

$$\mathcal{F}_{\mathcal{T}}^5 := \langle \underline{N_n^*} \mathcal{T} \underline{N_n} \underline{N_n^*} \mathbf{A} \rangle \left(\langle \underline{N_n^*} \mathcal{T} \underline{N_n} \underline{N_n^*} \mathbf{A} \rangle \right)^* \langle \underline{N_n^*} \mathcal{T} \underline{N_n} \underline{N_n^*} \mathbf{A} \rangle \rrbracket \Sigma_{\mathcal{T}}^*$$

$$\text{COND}_{\mathcal{T}}^5 := \{ w \in \Sigma_{\mathcal{T}}^* \mid \text{COMB}_n \mathbf{A} \in \mathcal{F}_{\mathcal{T}}^5(v) \text{ for every proper suffix } v \text{ of } w \text{ such that } v[1] = \langle \}$$

Condition 6 (tile colours match vertically). Let $\blacktriangledown_t := \{ t' \in \mathcal{T} \mid \text{top}(t') = \text{bottom}(t) \}$ be the set of tiles whose top colour matches the bottom colour of tile t . Define

$$\begin{aligned} \mathcal{F}_{\mathcal{T}}^6 := \sum_{t \in \mathcal{T}} \left(\right. & \left. \langle \underline{N_n^*} t \underline{N_n^*} \mathbf{A} \rangle \left(\langle \underline{N_n^*} \mathcal{T} \underline{N_n^*} \mathbf{A} \rangle \right)^* \rrbracket \cdot \right. \\ & \left. \llbracket \left(\langle \underline{N_n^*} \mathcal{T} \underline{N_n^*} \mathbf{A} \rangle \right)^* \langle \underline{N_n^*} \blacktriangledown_t \underline{N_n^*} \mathbf{A} \rangle \left(\langle \underline{N_n^*} \mathcal{T} \underline{N_n^*} \mathbf{A} \rangle \right)^* \rrbracket \Sigma_{\mathcal{T}}^* \right). \end{aligned}$$

The expression above was typeset in two lines only to highlight the correspondence between cells in two consecutive rows. Define the language $COND_{\mathcal{T}}^6$ as

$$COND_{\mathcal{T}}^6 := \left\{ w \in \Sigma_{\mathcal{T}}^* \mid COMB_n \mathbf{A} \in \mathcal{F}_{\mathcal{T}}^6(v) \text{ for every proper suffix } v \text{ of } w \text{ such that} \right. \\ \left. v[1] = \langle \text{ and } v[j] = \mathbb{I} \text{ for some } j \right\}$$

Intuitively, requiring \mathbb{I} to appear in v ensures that v does not begin in the last row.

Observe that $\mathcal{A}(\mathcal{F}_{\mathcal{T}}^4)$ and $\mathcal{A}(\mathcal{F}_{\mathcal{T}}^5)$ have a constant number of states, and $\mathcal{A}(\mathcal{F}_{\mathcal{T}}^6)$ has $O(|\mathcal{T}|)$ states. Define $L_{\mathcal{T}} := \mathbf{A} \bigcap_{1 \leq i \leq 6} COND_{\mathcal{T}}^i$. Let us recall:

Lemma 3.4. $L_{\mathcal{T}} \neq \emptyset \iff$ there exists a valid $(\mathcal{T}, t^{\setminus}, t^{\setminus})$ -tiling of width 2^n .

This lemma directly follows from:

Lemma 3.9. $L_{\mathcal{T}} = VALIDENC_{\mathcal{T}}$.

Proof. The inclusion $L_{\mathcal{T}} \supseteq VALIDENC_{\mathcal{T}}$ is trivial.

INCLUSION $L_{\mathcal{T}} \subseteq VALIDENC_{\mathcal{T}}$. Take any $u \in L_{\mathcal{T}}$. Due to Condition 1, it has the form $\mathbf{A} \prod_{1 \leq i \leq h} \mathbb{I} v_i \mathbb{I}$, for some $h \in \mathbb{N}$, where each $v_i \in (\Sigma_{\mathcal{T}} \setminus \{\mathbb{I}, \mathbb{I}\})^*$. We will show that $\mathbb{I} v_i \mathbb{I}$ is a proper encoding of a row under $encRow_{\mathcal{T}}$. Fix an arbitrary i . Again due to Condition 1, v_i has the form

$$\prod_{1 \leq j \leq w_i} (\langle p_{i,j} t_{i,j} s_{i,j} \mathbf{A} \rangle),$$

where $w_i \in \mathbb{N}$, $p_{i,j}, s_{i,j} \in N_n^*$, $p_{i,1} = s_{i,w_i} = n$, and $t_{i,j} \in \mathcal{T}$. Due to Condition 5, we have that all $s_{i,j}$ are nonempty and

$$p_{i,1} s_{i,1}[1] s_{i,2}[1] s_{i,3}[1] \cdots s_{i,w_i}[1] = COMB_n. \quad (3.4)$$

This implies that $w_i = 2^n$. By Condition 5 and (3.4) we get that $p_{i,j} = COMB_n[1, j]$, and now Condition 4 implies that $s_{i,j} = COMB_n[j + 1, 2^n + 1]$, so $\mathbb{I} v_i \mathbb{I}$ is a valid encoding of a row of length 2^n . Hence u encodes a tiling $T := [t_{i,j}]_{i,j} \in \mathcal{T}^{h \times 2^n}$. Property (3.2) in the definition of a valid tiling is now trivially implied by Condition 3, and we only need to show (3.1). Fix an arbitrary pair of tiles $t_{i,j}, t_{i+1,j}$ that are vertical neighbours. Observe that $p_{i,j} s_{i+1,x} = COMB_n \iff x = j$. Therefore, by Condition 6, $bottom(t_{i,j}) = top(t_{i+1,j})$, thus T is a valid \mathcal{T} -tiling, and, by Condition 2, a valid $(\mathcal{T}, t^{\setminus}, t^{\setminus})$ -tiling. \square

3.4. Construction of the automaton $\mathcal{A}_{\mathcal{T}}$. Let $\Sigma_{\mathcal{T},\#} := \Sigma_{\mathcal{T}} \cup \{\#\}$. Here, we define the NFA $\mathcal{A}_{\mathcal{T}}$ over $\Sigma_{\mathcal{T},\#}^2$ and prove Lemma 3.5, which states that $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{T}})) = L_{\mathcal{T}}$. The construction we present in this section, however, does not require the full generality of the setting of automatic structures:

- $Lang2Rel(\mathcal{L}(\mathcal{A}_{\mathcal{T}}))$ only holds for words of the same length, i.e., $\mathcal{A}_{\mathcal{T}}$ rejects words with $\#$;
- we only use a subset of the alphabet: $\Sigma_{\mathcal{T}} \times N_n \subset \Sigma_{\mathcal{T},\#}^2$.

For this reason, we begin with a lemma, which allows us to focus only on words satisfying the above properties. Let $\rho_{\mathcal{T}}: (\Sigma_{\mathcal{T}} \times N_n)^* \rightarrow \Sigma_{\mathcal{T}}^*$ be a homomorphism given by $\rho_{\mathcal{T}}(a, \cdot) := a$. Additionally, let

$$\rho_{\mathcal{T}}^{\forall}(L) := \{w \in \Sigma_{\mathcal{T}}^* \mid \rho_{\mathcal{T}}^{-1}(w) \subseteq L\}.$$

Lemma 3.10. For any NFA $\mathcal{A}'_{\mathcal{T}}$ over $\Sigma_{\mathcal{T}} \times N_n$ with $O(n \cdot |\mathcal{T}|)$ states, there exists an NFA $\mathcal{A}_{\mathcal{T}}$ over $\Sigma_{\mathcal{T},\#}^2$ with $O(n \cdot |\mathcal{T}|)$ states such that $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{T}})) = \rho_{\mathcal{T}}^{\forall}(\mathcal{L}(\mathcal{A}'_{\mathcal{T}}))$.

Proof. Take any $\mathcal{A}'_{\mathcal{I}}$ over $\Sigma_{\mathcal{I}} \times N_n$. Let

$$\begin{aligned} \mathcal{E}_1 &:= (\Sigma_{\mathcal{I}}^2)^*(\Sigma_{\mathcal{I}} \times \{\#\})^+ + (\Sigma_{\mathcal{I}}^2)^*(\{\#\} \times \Sigma_{\mathcal{I}})^+ && (u \otimes v \text{ such that } |u| \neq |v|) \\ \mathcal{E}_2 &:= (\Sigma_{\mathcal{I}}^2)^*(\Sigma_{\mathcal{I}} \times (\Sigma_{\mathcal{I}} \setminus N_n))(\Sigma_{\mathcal{I}}^2)^* && (\text{words with letter from } \Sigma_{\mathcal{I}}^2 \setminus (\Sigma_{\mathcal{I}} \times N_n)) \\ \mathcal{A}_{\mathcal{I}} &:= \mathcal{A}'_{\mathcal{I}} \oplus \mathcal{A}(\mathcal{E}_1) \oplus \mathcal{A}(\mathcal{E}_2). \end{aligned}$$

By definition, a word w belongs to $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{I}}))$ whenever for all v the word $w \otimes v$ belongs to $\mathcal{L}(\mathcal{A}_{\mathcal{I}})$. By construction, $\mathcal{L}(\mathcal{A}_{\mathcal{I}})$ contains all $w \otimes v$ where $|v| \neq |w|$ (\mathcal{E}_1) or where v is using a symbol from $\Sigma_{\mathcal{I}} \setminus N_n$ (\mathcal{E}_2). Hence, the only words which can be missing from $\mathcal{L}(\mathcal{A}_{\mathcal{I}})$ come from $\mathcal{L}(\mathcal{A}'_{\mathcal{I}})$. This implies that $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{\mathcal{I}})) = \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{A}'_{\mathcal{I}}))$. \square

Therefore, we only have to provide $\mathcal{A}'_{\mathcal{I}}$ such that $\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{A}'_{\mathcal{I}})) = L_{\mathcal{I}}$. The construction is modular, based on six NFA corresponding to Conditions 1 to 6:

Lemma 3.11 (modular design). *For any six NFA $(\mathcal{C}_{\mathcal{I}}^i)_{1 \leq i \leq 6}$ over $\Sigma_{\mathcal{I}} \times N_n$ with $O(n \cdot |\mathcal{T}|)$ states, there exists an NFA $\mathcal{A}'_{\mathcal{I}}$ with $O(n \cdot |\mathcal{T}|)$ states over $\Sigma_{\mathcal{I}} \times N_n$ such that*

$$\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{A}'_{\mathcal{I}})) = \mathbf{A} \bigcap_{1 \leq i \leq 6} \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_{\mathcal{I}}^i)).$$

Proof. Define

$$\begin{aligned} \mathcal{H} &:= (\{\mathbf{A}\} \times N_n \setminus \{1, 2, \dots, 6\}) (\Sigma_{\mathcal{I}} \times N_n)^* \\ \mathcal{A}'_{\mathcal{I}} &:= \mathcal{A}((\mathbf{A}, 1)) \odot \mathcal{C}_{\mathcal{I}}^1 \oplus \mathcal{A}((\mathbf{A}, 2)) \odot \mathcal{C}_{\mathcal{I}}^2 \oplus \dots \oplus \mathcal{A}((\mathbf{A}, 6)) \odot \mathcal{C}_{\mathcal{I}}^6 \oplus \mathcal{A}(\mathcal{H}). \end{aligned}$$

Observe that

$$\begin{aligned} \mathbf{A}w \in \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{A}'_{\mathcal{I}})) &\iff \rho_{\mathcal{I}}^{-1}(\mathbf{A}w) \subseteq \mathcal{L}(\mathcal{A}'_{\mathcal{I}}) \iff (\{\mathbf{A}\} \times N_n) \rho_{\mathcal{I}}^{-1}(w) \subseteq \mathcal{L}(\mathcal{A}'_{\mathcal{I}}) \iff \\ &\iff \forall i \in N_n. (\mathbf{A}, i) \rho_{\mathcal{I}}^{-1}(w) \subseteq \mathcal{L}(\mathcal{A}'_{\mathcal{I}}), \end{aligned}$$

but trivially

$$\begin{aligned} \mathcal{L}(\mathcal{A}((\mathbf{A}, j)) \odot \mathcal{C}_{\mathcal{I}}^j) \cap (\mathbf{A}, i) \rho_{\mathcal{I}}^{-1}(w) &= \emptyset && \text{for any } 1 \leq i, j \leq 6, i \neq j \\ \mathcal{L}(\mathcal{A}(\mathcal{H})) \cap (\mathbf{A}, i) \rho_{\mathcal{I}}^{-1}(w) &= \emptyset && \text{for any } 1 \leq i \leq 6. \end{aligned}$$

Therefore, $\mathbf{A}w \in \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{A}'_{\mathcal{I}}))$ if, and only if, $\rho_{\mathcal{I}}^{-1}(w) \subseteq \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_{\mathcal{I}}^i))$ for all $1 \leq i \leq 6$. \square

By the definition of $L_{\mathcal{I}}$, it only remains to construct automata $\mathcal{C}_{\mathcal{I}}^i$ such that $\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_{\mathcal{I}}^i)) = \text{COND}_{\mathcal{I}}^i$ for $1 \leq i \leq 6$. The construction is easy for Conditions 1 to 3:

$$\mathcal{C}_{\mathcal{I}}^i := \rho_{\mathcal{I}}^{-1}(\mathcal{B}_{\mathcal{I}}^i) \quad \text{for } i \in \{1, 2, 3\}$$

as $\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\rho_{\mathcal{I}}^{-1}(\mathcal{A}))) = \mathcal{L}(\mathcal{A})$ for any NFA \mathcal{A} . Observe that the number of states of all $\mathcal{C}_{\mathcal{I}}^i$ is upper bound by $O(\mathcal{T})$. The remaining Conditions 4 to 6 all talk about “every proper suffix” satisfying some simple regular property. We handle that in a general way. For $L \subseteq (\Sigma_{\mathcal{I}} \times N_n)^*$, define

$$L_{\forall \text{suf}}(L) := \left\{ w \mid v \in \rho_{\mathcal{I}}^{\forall}(L) \text{ for all proper suffixes } v \text{ of } w \right\}$$

Lemma 3.12 (recognising “for all proper suffixes”). *For any NFA \mathcal{A} over $\Sigma_{\mathcal{I}} \times N_n$ with k states, there exists an NFA $\text{ALLSUF}(\mathcal{A})$ with $O(k)$ states such that*

$$\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\text{ALLSUF}(\mathcal{A}))) = L_{\forall \text{suf}}(\mathcal{L}(\mathcal{A})).$$

Proof. Fix any NFA $\mathcal{A} = (Q, \Sigma_{\mathcal{I}} \times N_n, \delta, Q_{\mathcal{I}}, Q_{\mathcal{F}})$. We define $ALLSUF(\mathcal{A})$ which guesses the suffix to verify

$$ALLSUF(\mathcal{A}) := (Q \cup \{s\}, \Sigma_{\mathcal{I}} \times N_n, \delta \cup \delta', \{s\}, Q_{\mathcal{F}} \cup \{s\})$$

for some fresh state $s \notin Q$, and δ' containing transitions $s \xrightarrow{(a,0)} s$ for $a \in \Sigma_{\mathcal{I}}$ and $s \xrightarrow{(a,i)} q$ for $a \in \Sigma_{\mathcal{I}}$, $i \in N_n^{>0}$, $q \in Q_{\mathcal{I}}$. Also, let τ be the homomorphism such that $\tau(a) := (a, 0)$.

INCLUSION “ \subseteq ”. Take any $w \in \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(ALLSUF(\mathcal{A})))$. Let v be any proper suffix of w . Take any $v' \in \rho_{\mathcal{I}}^{-1}(v)$. We need to show that $v' \in \mathcal{L}(\mathcal{A})$. The word w can be written as uav , for $|u| \geq 0$ and $|a| = 1$. Consider a word $w' = \tau(u)(a, 1)v'$. By definition of $\rho_{\mathcal{I}}^{\forall}$, $w' \in \mathcal{L}(ALLSUF(\mathcal{A}))$. Let r be an accepting run of $ALLSUF(\mathcal{A})$ over w' . By construction, the run stays in state s while reading $\tau(u)$ and goes to some $q \in Q_{\mathcal{I}}$ upon reading $(a, 1)$. Therefore, the remaining suffix of r is an accepting run of \mathcal{A} over v' .

INCLUSION “ \supseteq ”. Fix $w \in L_{\forall \text{suf}}(\mathcal{L}(\mathcal{A}))$. Take any $w' \in \rho_{\mathcal{I}}^{-1}(w)$. We will show that $w' \in \mathcal{L}(ALLSUF(\mathcal{A}))$. Let $u'(a, k)v' := w'$ be such that u' is the maximal prefix arising as $\tau(u)$ for some u (possibly empty). Note that $k \neq 0$. By assumption, $v' \in \mathcal{L}(\mathcal{A})$, so there exists an accepting run r_2 of \mathcal{A} over v' starting in some $q \in Q_{\mathcal{I}}$. By construction, there exists a run r_1 from s to q over $u'(a, k)$ in $ALLSUF(\mathcal{A})$. Hence, the run $r_1 r_2$ accepts w' . \square

To handle conditions “beginning with \langle ” and “containing \mathbb{I} ” appearing as antecedents of implications, we proceed in the vein of the equivalence $a \rightarrow b \equiv \neg a \vee b$. Let

$$\mathcal{G}_{-\langle} := (\Sigma_{\mathcal{I}} \setminus \{\langle\}) \Sigma_{\mathcal{I}}^* \qquad \mathcal{G}_{-\mathbb{I}} := (\Sigma_{\mathcal{I}} \setminus \{\mathbb{I}\})^*.$$

Lemma 3.13. *For any $i \in \{4, 5, 6\}$, given an NFA $\hat{\mathcal{C}}_{\mathcal{I}}^i$ of size $O(n \cdot |\mathcal{T}|)$ and satisfying*

$$\rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^i)) = \{w \mid COMB_n \mathbf{A} \in \mathcal{F}_{\mathcal{I}}^i(w)\},$$

one can construct $\mathcal{C}_{\mathcal{I}}^i$ with $O(n \cdot |\mathcal{T}|)$ states such that $\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_{\mathcal{I}}^i)) = COND_{\mathcal{I}}^i$.

Proof. Fix $\hat{\mathcal{C}}_{\mathcal{I}}^4, \hat{\mathcal{C}}_{\mathcal{I}}^5, \hat{\mathcal{C}}_{\mathcal{I}}^6$ as in the statement of the lemma. We define $\mathcal{C}_{\mathcal{I}}^i$ as

$$\begin{aligned} \mathcal{C}_{\mathcal{I}}^4 &:= ALLSUF(\hat{\mathcal{C}}_{\mathcal{I}}^4 \oplus \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{G}_{-\langle}))) \\ \mathcal{C}_{\mathcal{I}}^5 &:= ALLSUF(\hat{\mathcal{C}}_{\mathcal{I}}^5 \oplus \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{G}_{-\langle}))) \\ \mathcal{C}_{\mathcal{I}}^6 &:= ALLSUF(\hat{\mathcal{C}}_{\mathcal{I}}^6 \oplus \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{G}_{-\langle} + \mathcal{G}_{-\mathbb{I}}))). \end{aligned}$$

The above cases are similar; w.l.o.g. let us focus on $\mathcal{C}_{\mathcal{I}}^4$. Observe that

$$\rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^4 \oplus \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{G}_{-\langle})))) = \rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^4)) \cup \mathcal{L}(\mathcal{G}_{-\langle}) = \{w \mid COMB_n \mathbf{A} \in \mathcal{F}_{\mathcal{I}}^4(w)\} \cup \mathcal{L}(\mathcal{G}_{-\langle}),$$

which directly corresponds to Condition 4, as required. Observe that the arguments to $ALLSUF(\cdot)$ are NFA with $O(n \cdot |\mathcal{T}|)$ states. \square

The essential element needed to define NFA $\hat{\mathcal{C}}_{\mathcal{I}}^i$ as in Lemma 3.13 is an NFA for the language $\{COMB_n \mathbf{A}\}$. First, we define $COMB_n$ as an intersection of languages of $n+1$ regular expressions. We then show how that intersection can be concisely represented by an NFA \mathcal{C}_n with $O(n)$ states such that $\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_n)) = \{COMB_n \mathbf{A}\}$.

Definition 3.14. We define $n + 1$ regular expressions \mathcal{E}_i over $\Sigma_{\mathcal{I}}$

$$\begin{aligned}\mathcal{E}_0 &:= N_n^{>1} (0 N_n^{>1})^* \\ \mathcal{E}_i &:= N_n^{>i} \left((N_n^{<i})^* i (N_n^{<i})^* N_n^{>i} \right)^* \quad \text{for } 0 < i < n \\ \mathcal{E}_n &:= n (N_n^{<n})^* n.\end{aligned}$$

Lemma 3.15. $\{COMB_n\} = \bigcap_{0 \leq i \leq n} \mathcal{L}(\mathcal{E}_i)$.

Proof. It is easy to prove the inclusion “ \subseteq ” by unravelling the definition of $COMB_n$. INCLUSION “ \supseteq ”. Take any $w \in \bigcap_{0 \leq i \leq n} \mathcal{L}(\mathcal{E}_i)$. We will show that $w = COMB_n$.

Claim 3.16. For $0 \leq k \leq n - 1$, we have $\bigcap_{0 \leq i \leq k} \mathcal{L}(\mathcal{E}_i) = \mathcal{L}\left(N_n^{>k} (COMB'_k N_n^{>k})^*\right)$.

We prove the claim by induction. The base case is trivial. Fix a word

$$w \in \mathcal{L}\left(N_n^{>k} \left(COMB'_k N_n^{>k}\right)^*\right) \cap \mathcal{L}(\mathcal{E}_{k+1}).$$

It has the form $w = a_1 COMB'_k a_2 COMB'_k \cdots COMB'_k a_m$ for some $m \geq 2$ and $a_1, a_2, \dots, a_m \in N_n^{>k}$. But since $w \in \mathcal{L}(\mathcal{E}_{k+1})$, all of a_2, a_4, a_6, \dots are equal to $k + 1$, and m is odd. Thus

$$w = a_1 \underbrace{COMB'_k (k+1) COMB'_k}_{COMB'_{k+1}} \cdots \underbrace{COMB'_k (k+1) COMB'_k}_{COMB'_{k+1}} a_m$$

We conclude by noticing that $\mathcal{L}(\mathcal{E}_n) \cap \mathcal{L}\left(N_n^{>(n-1)} \left(COMB'_{n-1} N_n^{>(n-1)}\right)^*\right) = \{COMB_n\}$. \square

Let us define

$$\mathcal{C}_n := \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{E}_0)) \odot \mathcal{A}((\mathbf{A}, 0)) \oplus \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{E}_1)) \odot \mathcal{A}((\mathbf{A}, 1)) \oplus \cdots \oplus \rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{E}_n)) \odot \mathcal{A}((\mathbf{A}, n)).$$

Observe that each $\mathcal{A}(\mathcal{E}_i)$ and $\mathcal{A}((\mathbf{A}, i))$ has a constant number of states, hence by further applying Fact 2.3, \mathcal{C}_n has $O(n)$ states.

Lemma 3.17. $\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_n)) = \left(\bigcap_{0 \leq i \leq n} \mathcal{L}(\mathcal{E}_i)\right) \mathbf{A}$.

Proof. INCLUSION “ \subseteq ”. Take any $w = u\mathbf{A} \in \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_n))$, and $i \in N_n$. We prove that $u \in \mathcal{L}(\mathcal{E}_i)$. By definition, $\rho_{\mathcal{I}}^{-1}(u\mathbf{A}) \subseteq \mathcal{L}(\mathcal{C}_n)$. Fix a homomorphism $\tau(a) = (a, 0)$. Note that $\tau(u)(\mathbf{A}, i) \in \mathcal{L}(\mathcal{C}_n)$. This can be accepted only by the $\rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{E}_i)) \odot \mathcal{A}((\mathbf{A}, i))$ component, thus $u \in \mathcal{L}(\mathcal{A}(\mathcal{E}_i)) = \mathcal{L}(\mathcal{E}_i)$, as required.

INCLUSION “ \supseteq ”. Take any $w = u\mathbf{A} \in \left(\bigcap_{0 \leq i \leq n} \mathcal{L}(\mathcal{E}_i)\right) \mathbf{A}$. Take any $u'(\mathbf{A}, i) \in \rho_{\mathcal{I}}^{-1}(u\mathbf{A})$. Since $u \in \mathcal{L}(\mathcal{E}_i)$, $u' \in \rho_{\mathcal{I}}^{-1}(\mathcal{L}(\mathcal{E}_i))$, and $u'(\mathbf{A}, i) \in \mathcal{L}(\rho_{\mathcal{I}}^{-1}(\mathcal{A}(\mathcal{E}_i)) \odot \mathcal{A}((\mathbf{A}, i)))$, as required. \square

Definition 3.18 (NFA $\hat{\mathcal{C}}_{\mathcal{I}}^i$). Let \mathcal{I} be of width 2^n and let $i \in \{4, 5, 6\}$. Take the NFA $\mathcal{C}_n = (Q^{(1)}, \Sigma_{\mathcal{I}} \times N_n, \delta^{(1)}, Q_{\mathbf{I}}^{(1)}, Q_{\mathbf{F}}^{(1)})$ as above with $O(n)$ states and $\mathcal{A}(\mathcal{F}_{\mathcal{I}}^i) = (Q^{(2)}, \Sigma_{\mathcal{I}} \times \Phi, \delta^{(2)}, Q_{\mathbf{I}}^{(2)}, Q_{\mathbf{F}}^{(2)})$ with $O(|\mathcal{T}|)$ states. We define $\hat{\mathcal{C}}_{\mathcal{I}}^i := (Q, \Sigma_{\mathcal{I}} \times N_n, \delta, Q_{\mathbf{I}}, Q_{\mathbf{F}})$ with $O(n \cdot |\mathcal{T}|)$ states, where

$$Q := Q^{(1)} \times Q^{(2)}, \quad Q_{\mathbf{I}} := Q_{\mathbf{I}}^{(1)} \times Q_{\mathbf{I}}^{(2)}, \quad Q_{\mathbf{F}} := Q_{\mathbf{F}}^{(1)} \times Q_{\mathbf{F}}^{(2)},$$

and the transition relation is

$$\delta := \left\{ (p, q) \xrightarrow{(a, \alpha)} (r, s) \mid q \xrightarrow{(a, \top)} s \in \delta^{(2)} \wedge p \xrightarrow{(a, \alpha)} r \in \delta^{(1)} \right\} \cup \left\{ (p, q) \xrightarrow{(a, \alpha)} (p, s) \mid q \xrightarrow{(a, \perp)} s \in \delta^{(2)} \wedge p \in Q^{(1)} \right\}.$$

Intuitively, $\hat{\mathcal{C}}_{\mathcal{I}}^i$ runs \mathcal{C}_n over the fragments of the input which are underlined by $\mathcal{F}_{\mathcal{I}}^i$.

Fact 3.19. We have $w \in \mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^i)$ if, and only if, there is some $v \in \mathcal{L}(\rho_{\mathcal{I}}^{-1}(\mathcal{F}_{\mathcal{I}}^i))$ such that $\psi_{\text{in}}(v) = w$ and $\psi_{\text{out}}(v) \in \mathcal{L}(\mathcal{C}_n)$.

To finish the construction, we need to prove the following property:

Lemma 3.20. For $i \in \{4, 5, 6\}$

$$\rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^i)) = \{w \mid \text{COMB}_n \mathbf{A} \in \mathcal{F}_{\mathcal{I}}^i(w)\}.$$

As the proofs for $i \in \{4, 5, 6\}$ are analogous, we focus on the hardest one, and then only comment how it can be adapted for $i \in \{4, 5\}$.

Proof ($i = 6$). PART A: INCLUSION “ \subseteq ”. Let $w \in \rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^6))$. Define

$$U := \{u \in \mathcal{L}(\mathcal{F}_{\mathcal{I}}^6) \mid \psi_{\text{in}}(u) = w\} \subseteq (\Sigma_{\mathcal{I}} \times \Phi)^*.$$

Note that if $U = \emptyset$, then $\mathcal{F}_{\mathcal{I}}^6(w) = \emptyset$. Hence, by Fact 3.19, $\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^6) = \emptyset$, and $\rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^6)) = \emptyset$, a contradiction. Therefore, $U \neq \emptyset$, and $w \in \mathcal{L}(\psi_{\text{in}}(\mathcal{F}_{\mathcal{I}}^6))$, so it is of the form

$$\langle p t s \mathbf{A} \rangle \beta \mathbb{I} \langle p_1 t_1 s_1 \mathbf{A} \rangle \langle p_2 t_2 s_2 \mathbf{A} \rangle \cdots \langle p_k t_k s_k \mathbf{A} \rangle \mathbb{I} \gamma$$

for some $k \in \mathbb{N}$, $p, p_i, s, s_i \in N_n^*$, $t, t_i \in \mathcal{T}$, $\beta \in (\Sigma_{\mathcal{I}} \setminus \{\mathbb{I}, \mathbb{I}\})^*$ and $\gamma \in \Sigma_{\mathcal{I}}^*$. Furthermore, $|U| = k$ and U contains the following underlined words $u_1^{(6)}, \dots, u_k^{(6)} \in (\Sigma_{\mathcal{I}} \times \Phi)^*$:

$$\begin{aligned} u_1^{(6)} &= \langle \underline{p} t s \mathbf{A} \rangle \beta \mathbb{I} \langle p_1 t_1 \underline{s_1} \mathbf{A} \rangle \langle p_2 t_2 s_2 \mathbf{A} \rangle \cdots \langle p_k t_k s_k \mathbf{A} \rangle \mathbb{I} \gamma \\ u_2^{(6)} &= \langle \underline{p} t s \mathbf{A} \rangle \beta \mathbb{I} \langle p_1 t_1 s_1 \mathbf{A} \rangle \langle p_2 t_2 \underline{s_2} \mathbf{A} \rangle \cdots \langle p_k t_k s_k \mathbf{A} \rangle \mathbb{I} \gamma \\ &\vdots \\ u_k^{(6)} &= \langle \underline{p} t s \mathbf{A} \rangle \beta \mathbb{I} \langle p_1 t_1 s_1 \mathbf{A} \rangle \langle p_2 t_2 s_2 \mathbf{A} \rangle \cdots \langle p_k t_k \underline{s_k} \mathbf{A} \rangle \mathbb{I} \gamma. \end{aligned}$$

We distinguish two cases based on whether some $u \in U$ satisfies $\psi_{\text{out}}(\rho_{\mathcal{I}}^{-1}(u)) \subseteq \mathcal{L}(\mathcal{C}_n)$.

CASE A.1: SUCH A u EXISTS. Take any such $u \in U$. Observe that $\psi_{\text{out}}(u) \in \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_n)) = \{\text{COMB}_n \mathbf{A}\}$. Hence, $\psi_{\text{in}}(u) = w$, $\psi_{\text{out}}(u) = \text{COMB}_n \mathbf{A}$, and $u \in \mathcal{L}(\mathcal{F}_{\mathcal{I}}^6)$. Therefore, $\text{COMB}_n \mathbf{A} \in \mathcal{F}_{\mathcal{I}}^6(w)$, as required.

CASE A.2: SUCH A u DOES NOT EXIST. Therefore, for every $u \in U$, there exists $v_u \in \rho_{\mathcal{I}}^{-1}(u)$ such that $\psi_{\text{out}}(v_u) \notin \mathcal{L}(\mathcal{C}_n)$. Fix any family $(v_u)_{u \in U}$ of such words. Let α_u be the position of the last underlined symbol in u . We now define a word $w' \in \rho_{\mathcal{I}}^{-1}(w)$ by choosing “problematic” output projections at the positions α_u . Formally, for every position i in w let

$$w'[i] := \begin{cases} \psi_{\text{out}}(v_u[i]) & \text{if } i = \alpha_u \text{ for some } u \\ (w[i], 0) & \text{otherwise.} \end{cases}$$

Observe that w' is properly defined, as positions α_u are pairwise different (corresponding to the letters \mathbf{A} following blocks s_1 to s_k). Since $\rho_{\mathcal{I}}(w') = w$ and $w \in \rho^{\forall}(\mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^6))$, we conclude

that $w' \in \mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^6)$. By Fact 3.19, we obtain $v \in \mathcal{L}(\rho_{\mathcal{I}}^{-1}(\mathcal{F}_{\mathcal{I}}^6))$ such that $\psi_{\text{in}}(v) = w'$ and $\psi_{\text{out}}(v) \in \mathcal{L}(\mathcal{C}_n)$. However, $\rho_{\mathcal{I}}(\psi_{\text{out}}(v)) = \rho_{\mathcal{I}}(\psi_{\text{out}}(v_u))$ for some $u \in U$ and the last symbols of $\psi_{\text{out}}(v)$ and $\psi_{\text{out}}(v_u)$ are identical. Since by construction \mathcal{C}_n ignores the component N_n of its alphabet $\Sigma_{\mathcal{I}} \times N_n$ for all letters but the last one, we get that

$$\psi_{\text{out}}(v) \in \mathcal{L}(\mathcal{C}_n) \quad \text{if and only if} \quad \psi_{\text{out}}(v_u) \in \mathcal{L}(\mathcal{C}_n).$$

This contradicts the fact that $\psi_{\text{out}}(v) \in \mathcal{L}(\mathcal{C}_n)$ and completes the analysis of Part A.

PART B: INCLUSION “ \supseteq ”. Consider w such that $\text{COMB}_n \mathbf{A} \in \mathcal{F}_{\mathcal{I}}^6(w)$. From the definition of $\mathcal{F}_{\mathcal{I}}^6(w)$, we can fix $v \in \mathcal{L}(\mathcal{F}_{\mathcal{I}}^6)$ such that $\psi_{\text{in}}(v) = w$ and $\psi_{\text{out}}(v) = \text{COMB}_n \mathbf{A}$. We need to show that $\rho_{\mathcal{I}}^{-1}(w) \subseteq \mathcal{L}(\hat{\mathcal{C}}_{\mathcal{I}}^6)$. Take any $w' \in \rho_{\mathcal{I}}^{-1}(w)$. By Fact 3.19, this requires constructing $u \in \mathcal{L}(\rho_{\mathcal{I}}^{-1}(\mathcal{F}_{\mathcal{I}}^6))$ such that $\psi_{\text{in}}(u) = w'$ and $\psi_{\text{out}}(u) \in \mathcal{L}(\mathcal{C}_n)$. We do this by combining w' and v . Let $u \in (\Sigma_{\mathcal{I}} \times N_n \times \Phi)^*$ be the unique word such that $\psi_{\text{in}}(u) = w'$ and for every position i in u we have $u[i] = (-, -, \top)$ if and only if $v[i] = (-, \top)$. We have to show that $\psi_{\text{out}}(u) \in \mathcal{L}(\mathcal{C}_n)$. Note that

$$\psi_{\text{out}}(u) \in \rho_{\mathcal{I}}^{-1}(\text{COMB}_n \mathbf{A}) = \rho_{\mathcal{I}}^{-1}\left(\left(\bigcap_{0 \leq i \leq n} \mathcal{L}(\mathcal{E}_i)\right) \mathbf{A}\right) = \rho_{\mathcal{I}}^{-1}\left(\rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_n))\right),$$

where the first inclusion follows from the definition of u , while the subsequent equalities follow from Lemmas 3.15 and 3.17, respectively. We get that $\rho_{\mathcal{I}}(\psi_{\text{out}}(u)) \in \rho_{\mathcal{I}}^{\forall}(\mathcal{L}(\mathcal{C}_n))$, and thus $\psi_{\text{out}}(u) \in \rho_{\mathcal{I}}^{-1}(\rho_{\mathcal{I}}(\psi_{\text{out}}(u))) \subseteq \mathcal{L}(\mathcal{C}_n)$. \square

Proof ($i \in \{4, 5\}$). The proof is analogous to the case $i = 6$. As the cases are distinguished by the filter $\mathcal{F}_{\mathcal{I}}^i$ being used, the only differences are related to the shape of words matched by $\psi_{\text{in}}(\mathcal{F}_{\mathcal{I}}^i)$. In particular, the set U for $i \in \{4, 5\}$ is now a singleton containing $u^{(i)}$:

$$u^{(4)} = \langle \underline{p} \underline{t} \underline{s} \mathbf{A} \rangle \gamma \quad (i = 4)$$

$$u^{(5)} = \langle \underline{p}_1 \underline{t} \underline{s}'_1 \underline{s}_1 \mathbf{A} \rangle \langle \underline{p}_2 \underline{t} \underline{s}'_2 \underline{s}_2 \mathbf{A} \rangle \cdots \langle \underline{p}_{k-1} \underline{t} \underline{s}'_{k-1} \underline{s}_{k-1} \mathbf{A} \rangle \langle \underline{p}_k \underline{t} \underline{s}'_k \mathbf{A} \rangle \parallel \gamma \quad (i = 5)$$

The rest of the proof only requires substituting $\mathcal{F}_{\mathcal{I}}^6$ with $\mathcal{F}_{\mathcal{I}}^4$ or $\mathcal{F}_{\mathcal{I}}^5$. \square

4. NFA OF DOUBLY EXPONENTIAL SIZE AFTER UNIVERSAL PROJECTION

From the lower bounds established in Section 3.4, it is now easy to construct a family $(\mathcal{A}_n)_{n \in \mathbb{N}}$ of NFA, each of size $O(n^4)$, such that the smallest NFA after a universal projection step has a doubly exponential number of states. This is achieved by simulating a binary counter over an exponential number of bits in a proper tiling.

To this end, for a fixed n , let \mathcal{I}_n be the problem instance from the left-hand side of Fig. 2, and define $\mathcal{A}_n := \mathcal{A}_{\mathcal{I}_n}$. Intuitively, the colours 0, 1 vertically represent the counter bits, and horizontally encode the carry-over bit. The only valid $(\mathcal{T}_{\text{inc}}, t^{\times}, t^{\times})$ -tiling of width $k > 2$ simulates incrementing an $(k - 2)$ -bit binary counter from 0 to $2^{k-2} - 1$; see the right-hand side of Figure 2 for an example for width $k = 5$. Thus, after a universal projection step, since \mathcal{A}_n enforces width $k = 2^n$, the resulting NFA accepts a single word of length doubly exponential in n .

Proposition 4.1. *The NFA for $\pi_1^{\forall}(\mathcal{L}(\mathcal{A}_{(\mathcal{T}_{\text{inc}}, t^{\times}, t^{\times}, n)}))$ has size $\Omega(2^{2^n})$.*

5. EMPTINESS AFTER UNIVERSAL PROJECTION IS IN EXPSPACE

We now consider algorithmic upper bounds for deciding whether the language of an automatic relation $R \subseteq (\Sigma^*)^{d+k}$, given by an NFA \mathcal{A}_R , is non-empty after a universal projection step. This yields the upper bound of Theorem 2.7.

Define a homomorphism $h: (\Sigma_{\#}^{d+k})^* \rightarrow (\Sigma_{\#}^d)^*$ by

$$h(a_1, \dots, a_d, a_{d+1}, \dots, a_{d+k}) := (a_1, \dots, a_d).$$

For a given NFA \mathcal{B} over $\Sigma_{\#}^{d+k}$ such that $S \subseteq (\Sigma^*)^{d+k}$ is automatic via \mathcal{B} (meaning that $S = \text{Lang2Rel}(\mathcal{L}(\mathcal{B}))$), it is clear that we can compute in linear time an NFA \mathcal{B}' with the same number of states as \mathcal{B} such that $L(\mathcal{B}') = h(\mathcal{L}(\mathcal{B}))$. The homomorphism h acts almost like existential projection, but in general, we do not have that $\pi_d^{\exists}(S)$ is automatic via \mathcal{B}' . For instance, suppose that

$$w = \begin{bmatrix} a \\ a \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} \begin{bmatrix} \# \\ c \end{bmatrix} \begin{bmatrix} \# \\ a \end{bmatrix} \in \mathcal{L}(\mathcal{B}).$$

Then $h(w) = aa\#\# \notin L_{\checkmark}$ because of the trailing $\#$ symbols. To remove them, we define

$$\text{STRIP}(L) := \left\{ w \mid \text{there exists } v \in (\{\#\}^d)^* \text{ such that } wv \in L \right\}.$$

It is then the case that $\pi_d^{\exists}(S)$ is automatic via $\text{STRIP}(\mathcal{L}(\mathcal{B}')) \cap L_{\checkmark}$. Note that an NFA for $\text{STRIP}(L)$ can be computed in linear time from an NFA for L without changing the set of states by making all states accepting that can reach a final state via a sequence of “ $\{\#\}^d$ ” symbols.

Recall that $\pi_d^{\forall}(R) = \overline{\pi_d^{\exists}(\overline{R})}$, an automatic presentation of $\pi_d^{\forall}(R)$ is thus given by

$$\overline{\left(\text{STRIP} \left(h \left(\overline{\mathcal{L}(\mathcal{A}_R)} \right) \right) \right) \cap L_{\checkmark}}.$$

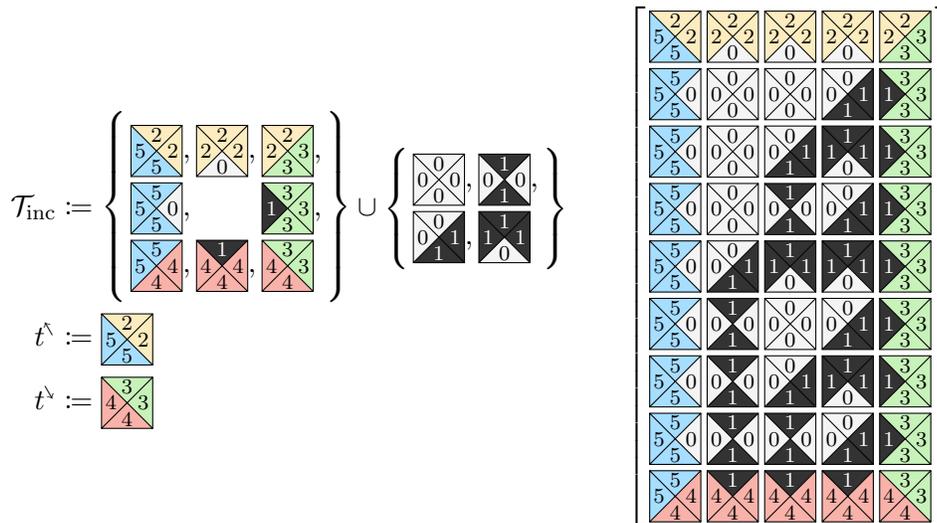


FIGURE 2. The instance of a tiling problem $\mathcal{I}_n = (\mathcal{T}_{\text{inc}}, t^x, t^y, n)$ and the unique valid tiling it enforces.

Assuming Q_R is the set of states of \mathcal{A}_R , and recalling that $L_{\checkmark} \subseteq (\Sigma_{\#}^d)^*$ is given by an NFA with 2^{d+2} states, it can easily be checked that the number of states of an NFA whose language gives the universal projection of R is bounded by $2^{((2^{|Q_R|+d+2})+d+2)}$.

With those characterisations and estimations at hand, the EXPSPACE upper bound stated in Theorem 2.7 can now easily be established.

Proposition 5.1. *Let R be an automatic relation given by an NFA \mathcal{A}_R . Deciding whether $\pi_d^{\forall}(R) \neq \emptyset$ is in EXPSPACE.*

Proof. For an EXPSPACE algorithm, we first construct an NFA $\mathcal{B} = (Q, \Sigma_{\#}^d, \delta, \{q_0\}, F)$ whose language is $(\text{STRIP}(h(\overline{\mathcal{L}(\mathcal{A}_R)})) \cap L_{\checkmark})$. We have $|Q| \leq 2^{|Q_R|+d+2}$, where Q_R is the set of states of \mathcal{A}_R , and hence \mathcal{B} can be constructed in exponential space. It remains to show that non-emptiness of $\overline{\mathcal{L}(\mathcal{B})} \cap L_{\checkmark}$ can be decided in polynomial space.

Note that we cannot explicitly construct an NFA for this language within polynomial space. Let $\mathcal{A}_{\checkmark} = (S, \Sigma_{\#}^d, \delta_{\checkmark}, \{s_0\}, F_{\checkmark})$ be the NFA for L_{\checkmark} , we can however non-deterministically guess a word in $\overline{\mathcal{L}(\mathcal{B})} \cap \mathcal{L}(\mathcal{A}_{\checkmark})$ letter by letter as follows. We keep track of a configuration of the form $(Q', s) \in 2^Q \times S$, which initially is $(\{q_0\}, s_0)$. Then we repeatedly non-deterministically guess some $a \in \Sigma_{\#}^d$ and update (Q', s) to $(\{\delta(q', a) \mid q' \in Q'\}, \delta_{\checkmark}(s, a))$ until we reach a configuration (Q', s) such that $Q' \cap F = \emptyset$ and $s \in F_{\checkmark}$. Clearly, the word obtained by this sequence of letters is in $\overline{\mathcal{L}(\mathcal{B})}$ and $\mathcal{L}(L_{\checkmark})$. The overall membership in EXPSPACE is then a consequence of Savitch's theorem and the observation that the length of the shortest word in $\overline{\mathcal{L}(\mathcal{B})} \cap L_{\checkmark}$ is bounded by $2^{((2^{|Q_R|+d+2})+d+2)}$. \square

6. LOWER BOUNDS FOR FRAGMENTS OF BÜCHI ARITHMETIC

In this section, we apply some of the ideas underlying the EXPSPACE lower bound for universal projection to develop some new lower bounds for fragments of Büchi arithmetic with a fixed quantifier alternation prefix. We explicitly state the lower bounds for Büchi arithmetic in base 2, they can, however, easily be generalised to any base $p \geq 2$. All reductions are easily seen to be LOGSPACE-computable and, again, we do not explicitly argue for that.

6.1. A lower bound for the $\exists^* \forall^* \exists^*$ fragment. We reduce from the *CORRIDORTILING* (Problem 3.2). Fix a problem instance $\mathcal{I} = (\mathcal{T}, t^{\wedge}, t^{\vee}, n)$ and let $W_n := 2^n$. The instance \mathcal{I} asks for the existence of a valid tiling of width W_n , with t^{\wedge} in the upper left corner and t^{\vee} in the bottom right one. W.l.o.g. we assume that $|\mathcal{T}| = 2^m$ for some $m \in \mathbb{N}$.

Let $\text{bin}(k) \in \{0, 1\}^+$ denote the binary representation of a number $k \in \mathbb{N}$ in most-significant-digit-first encoding such that $\text{bin}(k) \in 1 \cdot \{0, 1\}^*$ for all $k > 0$ and $\text{bin}(0) := 0$. In what follows, powers of 2 will be used to indicate positions in binary expansions of numbers. We use the predicate $\text{BitAt}(v, x, b)$ defined below for determining whether the bit b occurs in the expansion of some v at a position determined by x

$$\llbracket \text{BitAt} \rrbracket = \left\{ (v, x, b) \left| \begin{array}{l} x = 2^k \text{ and } v = \llbracket \text{pre} \cdot b \cdot \text{suf} \rrbracket_2 \text{ for some} \\ k \in \mathbb{N}, \text{pre} \in \{0, 1\}^*, \text{ and } \text{suf} \in \{0, 1\}^{k-1} \end{array} \right. \right\}.$$

This predicate can be implemented by an \exists^* formula of Büchi arithmetic as follows:

$$\text{BitAt}(v, x, b) := \exists pre \exists y \exists suf. \quad (6.1)$$

$$P_2(x) \wedge \quad (6.2)$$

$$(\bigvee_2(pre, y) \wedge y > x \vee pre = 0) \wedge \quad (6.3)$$

$$x > suf \wedge \quad (6.4)$$

$$((b = 0 \wedge v = pre + suf) \vee (b = 1 \wedge v = pre + x + suf)). \quad (6.5)$$

In words, we make sure that $\text{bin}(x)$ has just one bit set (6.2). Then, $\text{bin}(pre)$ is required to have all its digits 1 to the left of $(\log_2(x) + 1)$ th least significant bit, or no digits 1 at all (6.3); similarly, $\text{bin}(suf)$ has to have all digits 1 to the right (6.4). With all that, we can assert that $\text{bin}(x) = \text{bin}(pre) \cdot b \cdot \text{bin}(suf)$ (6.5).

For a given $k \in \mathbb{N}$, by using a $(k + 1)$ -tuple of numbers $\bar{u} = (u_0, \dots, u_k)$, we go beyond the binary alphabet $\{0, 1\}$ of $\text{bin}(\cdot)$ to the alphabet $\{0, 1, \dots, 2^{k+1} - 1\}$. The extra variable u_k in \bar{u} is for technical convenience in order to enable us to use the formulas defined in this section in the next section as well. We define an equivalent of BitAt :

$$\begin{aligned} \text{NumAt}_k(\bar{u}, x, a) &:= \exists a_0 \exists a_1 \dots \exists a_{k+1}. \exists b_1 \dots \exists b_{k+1}. \\ &a_0 = 0 \wedge a_{k+1} = a \wedge \\ &\bigwedge_{i=1}^{k+1} [a_i = 2a_{i-1} + b_i \wedge \text{BitAt}(u_{i-1}, x, b_i)]. \end{aligned}$$

Intuitively, bits at the position determined by x in \bar{u} are interpreted as the binary representation of a number. Note that NumAt_k is also an \exists^* formula.

Fix some enumeration $\psi: \{0, 1, \dots, 2^m - 1\} \rightarrow \mathcal{T}$ of tiles. We can trivially define quantifier-free formulas $\text{TopLeft}_{\mathcal{T}}$, $\text{BottomRight}_{\mathcal{T}}$, $\text{MatchH}_{\mathcal{T}}$, and $\text{MatchV}_{\mathcal{T}}$ such that

$$\begin{aligned} \llbracket \text{TopLeft}_{\mathcal{T}} \rrbracket &= \{t \in \mathbb{N} \mid \psi(t) = t^\wedge\} \\ \llbracket \text{BottomRight}_{\mathcal{T}} \rrbracket &= \{t \in \mathbb{N} \mid \psi(t) = t^\vee\} \\ \llbracket \text{MatchH}_{\mathcal{T}} \rrbracket &= \{(t, t') \in \mathbb{N}^2 \mid \text{right}(\psi(t)) = \text{left}(\psi(t'))\} \\ \llbracket \text{MatchV}_{\mathcal{T}} \rrbracket &= \{(t, t') \in \mathbb{N}^2 \mid \text{bottom}(\psi(t)) = \text{top}(\psi(t'))\}. \end{aligned}$$

Our goal is to represent a tiling of height H as a tuple of numbers \bar{u} such that the bottom-right tile is at position one, the top-left at position $W_n \cdot H$, and any tile t at position (x, y) is at position $W_n \cdot H - (y - 1) \cdot W_n - (x - 1)$. To assert that tiles are properly vertically aligned, we need to compare tiles whose distance is W_n . As stated above, in this section, $W_n = 2^n$ and hence of polynomial bit-length. In principle, we could now straight-forwardly continue with the reduction. However, in the next section we will have $W_n = 2^{2^n}$, which is of exponential bit length. In order to have a uniform presentation, in this section we give a slightly more sophisticated reduction that is here easy to follow and can then be enhanced to work for $W_n = 2^{2^n}$ in the next section.

Our first step is to define a “ruler”. Intuitively, $\text{IsARuler}_n(\bar{u}, F, W_n)$ holds if for every position p in the interval $[0, F)$, $\text{NumAt}_n(\bar{u}, 2^p, p \bmod W_n)$ holds:

$$\begin{aligned} \text{IsARuler}_n(\bar{u}, F, W) &:= \text{NumAt}_n(\bar{u}, 1, 0) \wedge \text{NumAt}_n(\bar{u}, 2F, 0) \wedge \\ &\forall x. (F > x \geq 1 \wedge P_2(x)) \rightarrow \\ &\exists v' \exists v. \text{NumAt}_n(\bar{u}, 2x, v') \wedge \text{NumAt}_n(\bar{u}, x, v) \wedge \\ &(v' = v + 1 \vee (v' = 0 \wedge v = W - 1)). \end{aligned}$$

Assuming that \bar{u} satisfies $\text{IsARuler}_n(\bar{u}, F, W)$, we can easily define the set of positions which are multiples of W_n and smaller than F :

$$\text{WidthMul}_n(\bar{u}, x) := \text{NumAt}_n(\bar{u}, x, 0).$$

Fact 6.1. We have $x = 2^k W_n$ for some $k \in \mathbb{N}$ if and only if for some F and \bar{u}

$$\text{IsARuler}_n(\bar{u}, F, W_n) \wedge 2F > x \geq 1 \wedge \text{WidthMul}_n(\bar{u}, x).$$

We can also ensure that positions represented by x and x' have distance exactly W_n :

$$\begin{aligned} \text{Width}_n(\bar{u}, x, x') := \exists v. \text{NumAt}_k(\bar{u}, x, v) \wedge \text{NumAt}_k(\bar{u}, x', v) \wedge \\ \forall y. (x > y > x') \rightarrow \neg \text{NumAt}_k(\bar{u}, y, v). \end{aligned}$$

Fact 6.2. We have $x = 2^{p+W_n}, y = 2^p$ for some p if and only if for some F and \bar{u}

$$\text{IsARuler}_n(\bar{u}, F, W_n) \wedge 2F > x \geq 1 \wedge \text{Width}_n(\bar{u}, x, y).$$

We now complete our reduction by defining a formula $\text{ETiling}_{\mathcal{I}}$ of Büchi arithmetic such that $\llbracket \text{ETiling}_{\mathcal{I}} \rrbracket$ is non-empty if and only if there exists a valid $(\mathcal{T}, t^\wedge, t^\vee)$ -tiling of width W_n :

$$\text{ETiling}_{\mathcal{I}} := \exists F. \exists u_0 \exists u_1 \dots \exists u_n. \exists w_1 \exists w_2 \dots \exists w_m.$$

$$\text{IsARuler}_n(\bar{u}, F, W_n) \wedge \tag{A}$$

$$[\exists t. \text{NumAt}_m(\bar{w}, F, t) \wedge \text{TopLeft}_{\mathcal{I}}(t)] \wedge \tag{B}$$

$$[\exists t. \text{NumAt}_m(\bar{w}, 1, t) \wedge \text{BottomRight}_{\mathcal{I}}(t)] \wedge \tag{C}$$

$$\begin{aligned} [\forall x. (F > x \wedge P_2(x) \wedge \neg \text{WidthMul}_n(\bar{u}, 2x, W_n)) \rightarrow \\ \exists t' \exists t. \text{NumAt}_m(\bar{w}, 2x, t') \wedge \text{NumAt}_m(\bar{w}, x, t) \wedge \text{MatchH}_{\mathcal{I}}(t', t)] \wedge \tag{D} \end{aligned}$$

$$\begin{aligned} [\forall x \forall x'. (2F > x > x' \wedge \text{Width}_n(\bar{u}, x, x', W_n)) \rightarrow \\ \exists t \exists t'. \text{NumAt}_m(\bar{w}, x, t) \wedge \text{NumAt}_m(\bar{w}, x', t') \wedge \text{MatchV}_{\mathcal{I}}(t, t')] \tag{E} \end{aligned}$$

Above, in (A), we require that $2F = 2^k W_n$ for some $k \in \mathbb{N}$. We interpret \bar{u} between positions F and 1 as a word over the alphabet $\{0, 1, \dots, 2^m - 1\}$. In (B) to (E), we express four properties of a valid tiling:

- in (B) and (C): that tiles t^\wedge and t^\vee are on their respective positions, F corresponding to the upper-left corner and 1 to the lower-right one,
- in (D): that tiles match horizontally; here, $2x$ and x are two consecutive powers of two, corresponding to horizontally adjacent tiles, unless $\text{WidthMul}_n(\bar{u}, 2x, N)$ holds, when $2x$ and x are in two consecutive rows,
- in (E): that tiles match vertically; positions x, x' correspond to vertical neighbours.

Keeping in mind that $\phi \rightarrow \psi \equiv \neg\phi \vee \psi$, we observe that $\text{ETiling}_{\mathcal{I}}$ has a quantifier prefix $\exists^* \forall^* \exists^*$ since WidthMul_n has an \exists^* quantifier prefix, Width_n has an $\exists^* \forall^*$ quantifier prefix, and IsARuler_n has an $\forall^* \exists^*$ quantifier prefix. Finally, observe that the above construction can easily be generalised to Büchi arithmetic with a V_p predicate for any $p \geq 2$. This yields:

Proposition 6.3. *Deciding the $\exists^* \forall^* \exists^*$ fragment of Büchi arithmetic is EXPSpace-hard.*

6.2. A lower bound for the $\exists^*\forall^*\exists^*\forall^*$ fragment. Building upon the reduction given in the previous section, we now give a 2-EXPSpace lower bound for the $\exists^*\forall^*\exists^*\forall^*$ fragment of Büchi arithmetic. We reduce from a modified variant of *CORRIDORTILING* (Problem 3.2), where we require the tiling width $W'_n := 2^{W_n} = 2^{2^n}$; this is trivially a 2-EXPSpace-hard problem since it allows for simulating a Turing machine with a tape with a doubly exponential number of cells. The reduction has the same general structure as that in Section 6.1, and the only difference lies in the definition of formulas WidthMul_n , Width_n , and IsARuler_n , which we substitute with $\text{WidthMul}'_n$, Width'_n , and $\text{IsARuler}'_n$, respectively, yielding a new formula $\text{ETiling}'_{\mathcal{L}}$. All three sub-formulas are given below. Observe that now $\text{ETiling}'_{\mathcal{L}}$ has the quantifier pattern $\exists^*\forall^*\exists^*\forall^*$ if we provide:

- $\text{WidthMul}'_n$ with an \exists^* quantifier prefix,
- Width'_n with an $\exists^*\forall^*\exists^*$ quantifier prefix,
- $\text{IsARuler}'_n$ with an $\exists^*\forall^*\exists^*\forall^*$ quantifier prefix.

We now give the definition of $\text{IsARuler}'_n$. Intuitively, the ruler in this case takes the shape of a comb from Section 3.3:

$$\begin{aligned} \text{IsARuler}'_n(\bar{u}, F, W) &:= \text{NumAt}_n(\bar{u}, 2F, W - 1) \wedge \text{NumAt}_n(\bar{u}, 1, W - 1) \wedge \\ &\quad \forall x \forall x' \forall v \forall v'. \\ &\quad (2F \geq x > x' \geq 1 \wedge \text{NumAt}_n(\bar{u}, x, v) \wedge \text{NumAt}_n(\bar{u}, x', v')) \rightarrow \\ &\quad [\exists y \exists v_y. x > y > x' \wedge \text{NumAt}_n(\bar{u}, y, v_y) \wedge v_y \geq \min\{v, v'\}] \vee \\ &\quad [\exists y \exists v_y. x > y > x' \wedge \text{NumAt}_n(\bar{u}, y, v_y) \wedge v_y = \min\{v, v'\} - 1 \wedge \\ &\quad \forall y' \forall v'_y. x < y' < x' \wedge y' \neq y \wedge \text{NumAt}_n(\bar{u}, y', v'_y) \rightarrow v'_y < v_y] \vee \\ &\quad [v = 0 \vee v' = 0] \end{aligned}$$

Let c be COMB_W without the first symbol, i.e., c has length W'_n . Intuitively, $\text{IsARuler}'_n$ defines a periodic sequence W_n, c^+ on \bar{u} observed through the lens of $\text{NumAt}_n(\bar{u}, x_i, \cdot)$ for consecutive positions x_i . We achieve this by requiring precisely one occurrence of the number $\min\{k, k'\} - 1$ between numbers $k, k' \in \mathbb{N}$, unless $\min\{k, k'\} = 0$ or there is some $k'' \geq \min\{k, k'\}$ between them. We can now simply put

$$\text{WidthMul}'_n(\bar{u}, x, W) := \text{NumAt}_n(\bar{u}, x, W).$$

Indeed, as the length of c is precisely W'_n and c contains only one occurrence of W_n at the rightmost position, positions marked with W_n correspond to $2^{\text{NW}'_n} \cap [1, 2F]$.

To verify that the distance between x and y is exactly W'_n , we first make sure that x and y fall into two consecutive rows using $\text{WidthMul}'_n$. If so, they are in the same position within a row, if the corresponding bits of the binary counter at x and y implicitly determined

by \bar{u} in line with Observation 3.7 are equal. This yields an $\exists^*\forall^*\exists^*$ formula:

$$\text{Width}'_n(\bar{u}, x, y, W) := \exists p \exists q \exists r. p > q > r \wedge \quad (\text{A.1})$$

$$\text{WidthMul}'_n(\bar{u}, p, W) \wedge \text{WidthMul}'_n(\bar{u}, q, W) \wedge \text{WidthMul}'_n(\bar{u}, r, W) \wedge \quad (\text{A.2})$$

$$[\forall s (p > s > r \wedge \text{WidthMul}'_n(\bar{u}, s, W)) \rightarrow s = q] \wedge \quad (\text{A.3})$$

$$p > x \geq q > y \geq r \wedge \quad (\text{A.4})$$

$$\forall B. (0 \leq B < W) \rightarrow \quad (\text{B})$$

$$[(\forall x'. (p > x' \geq x) \rightarrow \neg \text{NumAt}_k(\bar{u}, x', B)) \wedge \quad (\text{C.1})$$

$$(\forall y'. (q > y' \geq y) \rightarrow \neg \text{NumAt}_k(\bar{u}, y', B))] \vee \quad (\text{C.2})$$

$$\forall x' \forall y'. (\text{NumAt}_n(\bar{u}, x', B) \wedge \text{NumAt}_n(\bar{u}, y', B)) \rightarrow \quad (\text{D})$$

$$[\exists x''. x' > x'' \geq x \wedge \text{NumAt}_k(\bar{u}, x'', B)] \vee \quad (\text{E.1})$$

$$[\exists y''. y' > y'' \geq y \wedge \text{NumAt}_k(\bar{u}, y'', B)] \vee \quad (\text{E.2})$$

$$[(\forall x''. x' > x'' \geq x \wedge P_2(x) \rightarrow \exists v \text{NumAt}_k(\bar{u}, x'', v) \wedge v < B) \wedge \quad (\text{F.1})$$

$$(\forall y''. y' > y'' \geq y \wedge P_2(y) \rightarrow \exists v \text{NumAt}_k(\bar{u}, y'', v) \wedge v < B)] \vee \quad (\text{F.2})$$

$$[(\exists x'' \exists v. x' > x'' \geq x \wedge \text{NumAt}_k(\bar{u}, x'', v) \wedge v > B) \wedge \quad (\text{G.1})$$

$$(\exists y'' \exists v. y' > y'' \geq y \wedge \text{NumAt}_k(\bar{u}, y'', v) \wedge v > B)]. \quad (\text{G.2})$$

Intuitively, we assert that p, q, r delimit two consecutive rows in which x and y fall; cf. (A.1) to (A.4). The remainder of the formula, in (C.1) to (G.2), asserts that the B th bit of the (virtual) 2^n -bit counter is equal for x and y ; with (B), we perform that test for every B . In (C.1) and (C.2) we test whether B is absent in segments between p and x , and between q and y . If so, the B th bit is 0 for both x and y , and thus the B th bit equality test is successful. If it is the case only for one segment, x and y are not vertical neighbours, and the formula is false. If in turn both segments contain B , we catch the cases where x' or y' are not the rightmost occurrences of B in (E.1) and (E.2). If both x' and y' are rightmost occurrences, we require either both segments to consist only of values less than B ; cf. (F.1) and (F.2), or both to contain a value greater than B ; cf. (G.1) and (G.2). This makes sure that the value of the B th bit is equal for x and y , and therefore, they are vertical neighbours.

As before, the above construction can be generalised to Büchi arithmetic with a V_p predicate for any $p \geq 2$. This yields:

Proposition 6.4. *Deciding the $\exists^*\forall^*\exists^*$ fragment of Büchi arithmetic is 2-EXPSpace-hard.*

7. CONCLUSION

In this paper, we studied the computational complexity of eliminating universal quantifiers in automatic structures. We showed that, in general, this is a computationally challenging problem whose associated decision problem is EXPSpace-complete. Our result further reinforces the intuition already stemming from [Kus09] that, in general, the alternation of quantifiers requires “complex” automata. We also used the technical construction underlying the EXPSpace lower bound to obtain new lower bounds for deciding formulas of Büchi arithmetic with a fixed quantifier alternation prefix.

It would be interesting to understand whether it is possible to identify natural sufficient conditions on regular languages for which a universal projection step does not result in a doubly exponential blow-up and only leads to, e.g., polynomial or singly exponential growth. Results of this kind have been obtained in model-theoretic terms for structures of bounded degree [KL11, DH12], but we are not aware of a systematic study of questions of this kind on the level of regular languages.

Finally, the lower bounds established in Section 6 for deciding formulas of Büchi arithmetic with fixed quantifier alternation prefixes are not tight. It can be seen that the upper bounds, which can be derived from Section 5 for those classes of formulas, are off by one exponential. We leave it as an open problem of this paper to fully characterise the complexity of deciding Büchi arithmetic with an arbitrary but fixed quantifier alternation prefix. A first step could be to uniformise the constructions in Section 6 to enable showing k -EXPSpace-hardness for any $k \geq 1$. At present, it is not obvious to us how this could be achieved.

ACKNOWLEDGEMENTS

We would like to thank the anonymous CONCUR'23 and LMCS reviewers for their comments and suggestions, which helped us to improve the presentation of this paper. We are grateful to Alessio Mansutti for spotting a bug in Section 4. This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 852769, ARiAT).

REFERENCES

- [BFV23] Bernard Boigelot, Pascal Fontaine, and Baptiste Vergain. Universal first-order quantification over automata. In *Implementation and Application of Automata, CIAA*, Lect. Notes Comp. Sci., pages 91–102. Springer, 2023. doi:10.1007/978-3-031-40247-0_6.
- [BG00] Achim Blumensath and Erich Grädel. Automatic structures. In *Logic in Computer Science, LICS*, pages 51–62. IEEE Computer Society, 2000. doi:10.1109/LICS.2000.855755.
- [BHMV94] Véronique Bruyère, Georges Hansel, Christian Michaux, and Roger Villemaire. Logic and p -recognizable sets of integers. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2):191–238, 1994. doi:10.36045/bbms/1103408547.
- [Boi] Bernard Boigelot. The LASH toolset. <https://people.montefiore.uliege.be/boigelot/research/lash/index.html>.
- [Bru85] Véronique Bruyère. Entiers et automates finis. *Mémoire de fin d'études*, 1985.
- [Büc60] J. Richard Büchi. Weak second-order arithmetic and finite automata. *Math. Logic Quart.*, 6(1-6):66–92, 1960. doi:10.1002/malq.19600060105.
- [CH17] Dmitry Chistikov and Christoph Haase. On the complexity of quantified integer programming. In *International Colloquium on Automata, Languages, and Programming, ICALP*, volume 80 of *LIPICs*, pages 94:1–94:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPICs.ICALP.2017.94.
- [CHM22] Dmitry Chistikov, Christoph Haase, and Alessio Mansutti. Geometric decision procedures and the VC dimension of linear arithmetic theories. In *Logic in Computer Science, LICS*, pages 59:1–59:13. ACM, 2022. doi:10.1145/3531130.3533372.
- [DH12] Antoine Durand-Gasselin and Peter Habermehl. Ehrenfeucht-Fraïssé goes elementarily automatic for structures of bounded degree. In *Symposium on Theoretical Aspects of Computer Science, STACS*, volume 14 of *LIPICs*, pages 242–253. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2012. doi:10.4230/LIPICs.STACS.2012.242.

- [GHW19] Florent Guépin, Christoph Haase, and James Worrell. On the existential theories of Büchi arithmetic and linear p -adic fields. In *Logic in Computer Science, LICS*, pages 1–10. IEEE, 2019. doi:10.1109/LICS.2019.8785681.
- [GS66] Seymour Ginsburg and Edwin H. Spanier. Semigroups, Presburger formulas, and languages. *Pac. J. Math.*, 16(2):285 – 296, 1966.
- [Haa14] Christoph Haase. Subclasses of Presburger arithmetic and the weak EXP hierarchy. In *Computer Science Logic (CSL) and Logic in Computer Science (LICS), CSL-LICS*, pages 47:1–47:10. ACM, 2014. doi:10.1145/2603088.2603092.
- [Hod82] Bernard R. Hodgson. On direct products of automaton decidable theories. *Theor. Comput. Sci.*, 19(3):331 – 335, 1982. doi:10.1016/0304-3975(82)90042-1.
- [HR21] Christoph Haase and Jakub Rózycki. On the expressiveness of Büchi arithmetic. In *Foundations of Software Science and Computation Structures, FoSSaCS*, volume 12650 of *Lect. Notes Comp. Sci.*, pages 310–323. Springer, 2021. doi:10.1007/978-3-030-71995-1_16.
- [KL11] Dietrich Kuske and Markus Lohrey. Automatic structures of bounded degree revisited. *J. Symb. Log.*, 76(4):1352–1380, 2011. doi:10.2178/js1/1318338854.
- [KN95] Bakhadyr Khoussainov and Anil Nerode. Automatic presentations of structures. In *Logical and Computational Complexity, LCC*, volume 960 of *Lect. Notes Comp. Sci.*, pages 367–392. Springer, 1995. doi:10.1007/3-540-60178-3_93.
- [Kus09] Dietrich Kuske. Theories of automatic structures and their complexity. In *Conference on Algebraic Informatics, CAI*, volume 5725 of *Lect. Notes Comp. Sci.*, pages 81–98. Springer, 2009. doi:10.1007/978-3-642-03564-7_5.
- [LP09] Jérôme Leroux and Gérald Point. TaPAS: The Talence Presburger Arithmetic Suite. In *Tools and Algorithms for the Construction and Analysis of Systems, TACAS*, volume 5505 of *Lect. Notes Comp. Sci.*, pages 182–185. Springer, 2009. doi:10.1007/978-3-642-00768-2_18.
- [Mou16] Hamoon Mousavi. Automatic theorem proving in Walnut. *CoRR*, 2016. arXiv:1603.06017.
- [Pre29] Mojżesz Presburger. Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes Rendus du I congrès de Mathématiciens des Pays Slaves*, pages 92–101. 1929.
- [Sch16] Sylvain Schmitz. Complexity hierarchies beyond elementary. *ACM Trans. Comput. Theory*, 2016. doi:10.1145/2858784.
- [Sch19] François Schwarzentruber. The complexity of tiling problems, 2019. arXiv:1907.00102.
- [Tho68] Ken Thompson. Programming techniques: Regular expression search algorithm. *Commun. ACM*, 11(6):419–422, 1968. doi:10.1145/363347.363387.
- [vEB19] Peter van Emde Boas. The convenience of tilings. In *Complexity, Logic, and Recursion Theory*, pages 331–363. CRC Press, 2019. doi:10.1201/9780429187490-12.