On the Expressiveness of Büchi Arithmetic*

Christoph Haase\(^1\) and Jakub Różycki\(^2\)

\(^1\) Department of Computer Science, University of Oxford, UK
\(^2\) Institute of Informatics, University of Warsaw, Poland

Abstract. We show that the existential fragment of Büchi arithmetic is strictly less expressive than full Büchi arithmetic of any base, and moreover establish that its \(\Sigma_2\)-fragment is already expressively complete. Furthermore, we show that regular languages of polynomial growth are definable in the existential fragment of Büchi arithmetic.

1 Introduction

This paper studies the expressive power of Büchi arithmetic, which is an extension of Presburger arithmetic, the first-order theory of the structure \(\langle \mathbb{N}, 0, 1, + \rangle\). Büchi arithmetic additionally allows for expressing restricted divisibility properties while retaining decidability. 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 \(V_p(a, b)\) holds if and only if \(a\) is the largest power of \(p\) dividing \(b\) without remainder, i.e., \(a = p^k\), \(a \mid b\) and \(p \cdot a \not\mid b\).

Presburger arithmetic admits quantifier-elimination in the extended structure \(\langle \mathbb{N}, 0, 1, +, \{c \mid \cdot\}_{c > 1} \rangle\) additionally consisting of unary divisibility predicates \(c \mid \cdot\) for every \(c > 1\) [9]. It follows that the existential fragment of Presburger arithmetic is expressively complete, since any predicate \(c \mid \cdot\) can be expressed using an additional existentially quantified variable. We study the analogous question for Büchi arithmetic and show, as the main result of this paper, that its existential fragment is, in any base, strictly less expressive than full Büchi arithmetic. Notably, this result implies that a quantifier-elimination procedure à la Presburger cannot exist for Büchi arithmetic.

A central result about Büchi arithmetic is that it is an automatic structure: a set \(M \subseteq \mathbb{N}^n\) is definable in Büchi arithmetic of base \(p\) if and only if \(M\) is recognizable by a finite-state automaton under a base \(p\) encoding of the natural numbers. Equivalently, \(M\) is \(p\)-regular. This result was first stated by Büchi [3], albeit in an incorrect form, and later correctly stated and proved by Bruyère [1], see also [2]. Villemaire showed that the \(\Sigma_3\)-fragment of Büchi arithmetic is expressively complete [12, Cor. 2.4]. He established this result by showing how to

\* 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). Parts of this research were carried out while the first author was affiliated with the Department of Computer Science, University College London, UK.
construct a $\Sigma_2$-formula defining the language of a given finite-state automaton. We observe that Villemaire’s construction can actually be improved to a $\Sigma_2$-formula and thus obtain a full characterization of the expressive power of Büchi arithmetic in terms of the number of quantifier alternations.

Our approach to separating the expressiveness of existential Büchi arithmetic from full Büchi arithmetic in base $p$ is based on a counting argument. Given a set $M \subseteq \mathbb{N}$, define the counting function $d_M(n) := \#(M \cap \{p^{n-1}, \ldots, p^n - 1\})$ which counts the numbers of bit-length $n$ in base $p$ in $M$. If $M$ is definable in existential Büchi arithmetic of base $p$, we show that $d_M$ is either $O(n^c)$ for some $c \geq 0$, or at least $c \cdot p^n$ for some constant $c > 0$ and infinitely many $n \in \mathbb{N}$. Since, for instance, for $M_p \subseteq \mathbb{N}$ defined as the set of numbers with $p$-ary expansion in the regular language $(10 \mid 01)^*$, we have $d_M(n) \leq 2^{n/2}$ for all $n \in \mathbb{N}$ and $d_M(n) = 2^{n/2}$ for even $n \in \mathbb{N}$, and hence $M$ is not definable in existential Büchi arithmetic of base $p$. However, $M_p$, being $p$-regular implies that $M_p$ is definable by a $\Sigma_2$-formula of Büchi arithmetic of base $p$.

We also show that existential Büchi arithmetic defines all regular languages of polynomial density, encoded as sets of integers. Given a language $L$ has polynomial density, encoded as sets of integers. Given a language $L$, if moreover $L$ is regular then Szilard et al. [10] say that $L$ has polynomial density whenever $d_L(n)$ is $O(n^c)$ for some non-negative integer $c$. If moreover $L$ is regular then Szilard et al. show that $L$ is represented as a finite union of regular expressions of the form $v_0w_1^*v_1 \cdots w_k^*v_k$ such that $0 \leq k \leq c + 1$, $v_0, w_1, v_1, \ldots, v_k, w_k \in \Sigma^*$ [10, Thm. 3]. We show that existential Büchi arithmetic defines any language represented by a regular expression $v_0w_1^*v_1 \cdots w_k^*v_k$, which implies that existential Büchi arithmetic defines all regular languages of polynomial density.

## 2 Preliminaries

Given $v \in \mathbb{Z}^d$, we denote by $\|v\|_\infty$ the maximum norm of $v$. For a matrix $A \in \mathbb{Z}^{m \times d}$, we denote by $\|A\|_{1,\infty}$ the one-infinity norm of $A$.

Let $\Sigma$ be an alphabet and $w \in \Sigma^*$, we denote by $|w|$ the length of $w$. Given a set $U \subseteq \mathbb{N}$, we denote by $w^U := \{w^u : u \in U\}$. Thus, for example, $w^* = w^\mathbb{N}$.

For an integer $p \geq 2$, let $\Sigma_p := \{0, \ldots, p - 1\}$. We view words over $\Sigma_p$ as numbers encoded in $p$-ary most-significant bit first encoding. Tuples of numbers encoded in $p$-ary most-significant bit first encoding. Tuples of numbers encoded in $p$-ary most-significant bit first encoding. Tuples of numbers encoded in $p$-ary most-significant bit first encoding. For $w = v_m \cdots v_0 \in (\Sigma^p)^{m+1}$, we denote by $\lfloor w \rfloor_p \in \mathbb{N}^n$ the $n$-tuple

$$\lfloor w \rfloor_p := \sum_{i=0}^{m} v_i \cdot p^i.$$

Note that $\lfloor \cdot \rfloor_p$ is not injective since, e.g., 01 and 001 both encode the number one. Given $L \subseteq (\Sigma_p)^*$, we define

$$\lfloor L \rfloor_p := \{\lfloor w \rfloor_p : w \in L\} \subseteq \mathbb{N}^n.$$
Automata. A deterministic automaton is a tuple \( A = (Q, \Sigma, \delta, q_0, F) \), where

- \( Q \) is a set of states,
- \( \Sigma \) is a finite alphabet,
- \( \delta : Q \times \Sigma \rightarrow Q \cup \{ \perp \} \), where \( \perp \notin Q \), is the transition function,
- \( q_0 \in Q \) is the initial state, and
- \( F \subseteq Q \) is the set of final states.

For states \( q,r \in Q \) and \( u \in \Sigma \), we write \( q \xrightarrow{u} r \) if \( \delta(q,u) = r \), and extend \( \rightarrow \) inductively to words by stipulating, for \( w \in \Sigma^* \) and \( u \in \Sigma \), that \( q \xrightarrow{w,u} r \) if there is \( s \in Q \) such that \( q \xrightarrow{w} s \xrightarrow{u} r \). The language of \( A \) is defined as \( L(A) = \{ w \in \Sigma^* : q_0 \xrightarrow{w} q_f, q_f \in F \} \).

Note that \textit{a priori} we allow automata to have infinitely many states and to have partially defined transition functions (due to the presence of \( \perp \) in the co-domain of \( \delta \)). If \( Q \) is finite then we call \( A \) a deterministic finite automaton (DFA), and if in addition \( \Sigma = \Sigma^p \) for some \( p \geq 2 \) and \( n \geq 1 \) then \( A \) is called a \textit{p-automaton}. Throughout this paper, we assume, without loss of generality, that all states of a DFA are live, i.e., every state is reachable and can reach an accepting state.

Arithmetic theories. As stated in the introduction, Presburger arithmetic is the first-order theory of the structure \( \langle \mathbb{N}, 0, 1, + \rangle \), and Büchi arithmetic of base \( p \) the first-order theory of the extended structure \( \langle \mathbb{N}, 0, 1, +, V_p \rangle \). We write atomic formulas of Presburger arithmetic as \( a \cdot x = c \), where \( a = (a_1, \ldots, a_d)^T \) with \( a_i \in \mathbb{Z} \), \( c \in \mathbb{Z} \), and \( x = (x_1, \ldots, x_d) \) is a vector of unknowns. In Büchi arithmetic we additionally have atomic formulas \( V_p(x,y) \) for the unknowns \( x \) and \( y \). For technical convenience, we assert that \( V_p(x,0) \) never holds.\(^3\) We write \( \Phi(x) \) or \( \Phi(x) \) to indicate that \( x \) or a vector of unknowns \( x \) occurs free in \( \Phi \). If there are further free variables in \( \Phi \), we assume them to be implicitly existentially quantified.

We may without loss of generality assume that no negation symbol occurs in a formula of Büchi arithmetic. First, we have \( \neg(a \cdot x = c) \equiv a \cdot x \leq c - 1 \lor a \cdot x \geq c + 1 \), and the order relation \( \leq \) can easily be expressed by introducing an additionally existentially quantified variable. Moreover, \( \neg V_p(x,y) \) holds if and only if \( y = 0 \), \( x \) is not a power of \( p \), or \( x \) is not the largest power of \( p \) dividing \( y \) without remainder. Denote by \( P_p(x) \equiv V_p(x,x) \) the macro asserting that \( x \) is a power of \( p \), we have

\[
\neg P_p(x) \equiv \exists s \exists y \exists z : x = y + z \land V_p(y,x) \land V_p(s,z) \land y < s,
\]

and hence \( \neg V_p(x,y) \equiv y = 0 \lor \neg P_p(x) \lor \exists z : (V_p(z,y) \land \neg (x = z)) \).

Given a formula \( \Phi(x) \) of Büchi arithmetic of base \( p \), we define

\[
\llbracket \Phi(x) \rrbracket_p := \{ m \in \mathbb{N}^d : \Phi[m/x] \text{ is valid} \},
\]

\(^3\) Other conventions are possible, e.g., asserting that \( V_p(x,0) \) holds if and only if \( x = 1 \) as in \([2]\), but this does not change the sets of numbers definable in Büchi arithmetic.
where, for \( m = (m_1, \ldots, m_d) \) and \( x = (x_1, \ldots, x_d) \), \( \Phi[m/x] \) is the formula obtained from replacing every \( x_i \) by \( m_i \) in \( \Phi \). The set of sets of numbers definable in Presburger arithmetic is denoted by

\[
\text{PA} := \{ [\Phi(x)] : \Phi(x) \text{ is a formula of Presburger arithmetic} \}.
\]

Likewise, we define the sets of numbers definable in fragments of Büchi arithmetic of base \( p \) with a fixed number of quantifier-alternations as

\[
\Sigma_i\text{-BA}_p := \{ [\Phi(x)]_p : \Phi(x) \text{ is a } \Sigma_i\text{-formula of Büchi arithmetic of base } p \}.
\]

Finally, \( \text{BA}_p := \bigcup_{i \geq 1} \Sigma_i\text{-BA}_p \) denotes the sets of numbers definable in Büchi arithmetic of base \( p \).

For separating existential Büchi arithmetic from full Büchi arithmetic, we employ some tools from enumerative combinatorics. As defined in [14], a formula of parametric Presburger arithmetic with parameter \( t \) is a formula of Presburger arithmetic \( \Phi_t \) in which atomic formulas are of the form \( a \cdot x = c(t) \), where \( c(t) \) is a polynomial. For \( n \in \mathbb{N} \), we denote by \( \Phi_n \) the formula of Presburger arithmetic obtained from replacing \( c(t) \) in every atomic formula of \( \Phi_t \) by the value of \( c(n) \).

We associate to a formula \( \Phi_t(x) \) the counting function \( \#\Phi_t(x) : \mathbb{N} \rightarrow \mathbb{N} \) such that

\[
\#\Phi_t(x)(n) := \#\Phi_n(x).
\]

Throughout this paper, we constraint ourselves to formulas \( \Phi_t(x) \) of parametric Presburger arithmetic in which \( c(t) \) is the identity function and \( \#\Phi_t(x)(n) \) is finite for all \( n \in \mathbb{N} \).

**Definition 1.** A function \( f : \mathbb{N} \rightarrow \mathbb{N} \) is an eventual quasi-polynomial if there exist a threshold \( t \in \mathbb{N} \) and polynomials \( p_0, \ldots, p_{m-1} \in \mathbb{Q}[t] \) such that for all \( n > t \), \( f(n) = p_i(n) \) whenever \( n \equiv i \mod m \).

Given an eventual quasi-polynomial \( f \) with threshold \( t \) and \( n > t \), we denote by \( f_n \) the polynomial \( p_i \) such that \( n \equiv i \mod m \). We say that the polynomials \( p_0, \ldots, p_{m-1} \) constitute the eventual quasi-polynomial \( f \). A result by Woods [14, Thm. 3.5(b)] shows that the counting functions associated to parametric Presburger formulas are eventual quasi-polynomial.

**Proposition 1 (Woods).** Let \( \Phi_t(x) \) be a formula of parametric Presburger arithmetic. Then \( \#\Phi_t(x) \) is an eventual quasi-polynomial.

**Semi-linear sets.** A result by Ginsburg and Spanier establishes that the sets of numbers definable in Presburger arithmetic are semi-linear sets [6]. A linear set in dimension \( d \) is given by a base vector \( b \in \mathbb{N}^d \) and a finite set of period vectors \( P = \{ p_1, \ldots, p_n \} \subseteq \mathbb{N}^d \) and defines the set

\[
L(b, P) := \{ b + \lambda_1 \cdot p_1 + \cdots + \lambda_n \cdot p_n : \lambda_i \in \mathbb{N}, 1 \leq i \leq n \}.
\]

A semi-linear set is a finite union of linear sets. For a finite \( B \subseteq \mathbb{N}^d \), we write \( L(B, P) \) for \( \bigcup_{b \in B} L(b, P) \). Semi-linear sets of the form \( L(B, P) \) are called hybrid.
linear sets in [4], and it is known that the set of non-negative integer solutions of a system of linear Diophantine inequalities \( S: A \cdot x \geq c \) is a hybrid linear set [4].

Semi-linear sets in dimension one are also known as ultimately periodic sets.

In this paper, we represent an ultimately periodic set as a four-tuple \( U = (t, \ell, B, R) \), where \( t \geq 0 \) is a threshold, \( \ell > 0 \) is a period, \( B \subseteq \{0, \ldots, t - 1\} \) and \( R \subseteq \{0, \ldots, \ell - 1\} \), and \( U \) defines the set

\[
\|U\| := \bigcup_{b \in B} \{b\} \cup \{t + r + \ell \cdot i : r \in R, i \geq 0\}.
\]

3 The inexpressiveness of existential Büchi arithmetic

We now establish the main result of this paper and show that the existential fragment of Büchi arithmetic is strictly less expressive than general Büchi arithmetic.

**Theorem 1.** For any base \( p \), \( \Sigma_1 \text{-BA}_p \neq \text{BA}_p \). In particular, there exists a fixed regular language \( L \subseteq \{0, 1\}^* \) such that \( \|L\|_p \in \text{BA} \setminus \Sigma_1 \text{-BA}_p \) for every base \( p \).

Given a set \( M \subseteq \mathbb{N} \), recall that for a fixed base \( p \geq 2 \), \( d_M(n) \) counts the numbers of bit-length \( n \) in base \( p \) in \( M \). As already discussed in the introduction, we prove Theorem 1 by characterizing the growth of \( d_M \) for sets \( M \) definable in Büchi arithmetic.

For any formula \( \Phi(x) \) of existential Büchi arithmetic in prenex normal form, we can, with no loss of generality, assume that its matrix is in disjunctive normal form, i.e., a disjunction of systems of linear Diophantine equations with valuation constraints, each of the form

\[
A \cdot x = c \land \bigwedge_{i \in I} V_p(x_i, y_i),
\]

where the \( x_i \) and \( y_i \) are unknowns from the vector of unknowns \( x \). For \( M = \|\Phi(x)\|_p \), in order to determine the growth of \( d_M \), it suffices to determine the maximum growth occurring in any of its systems of linear Diophantine equations with valuation constraints in the matrix of \( \Phi(x) \), which in turn can be obtained by analyzing the growth of the number of words accepted by a \( p \)-automaton defining the set of solutions of such a system.

Let \( S: A \cdot x = c \) be a system of linear Diophantine equations such that, throughout this section, \( A \) is an \( m \times d \) integer matrix, and fix a base \( p \geq 2 \). Following Wolper and Boigelot [13], we define an automaton \( A(S) := (Q, \Sigma_d^d, 0, q_0, F) \) whose language encodes all solutions of \( S \) over the alphabet \( \Sigma_d^d \):

\[
\begin{align*}
- & Q := \mathbb{Z}^m, \\
- & \delta(q, u) := p \cdot q + A \cdot u \text{ for all } q \in Q \text{ and } u \in \Sigma_d^d, \\
- & q_0 := 0, \text{ and} \\
- & F := \{c\}.
\end{align*}
\]
As discussed in [13], see also [7], only states \( q \) such that \( \| q \|_\infty \leq \| A \|_{1,\infty} \) and \( \| q \|_\infty \leq \| c \|_\infty \) can reach the accepting state. Hence, all words \( w \in (\Sigma_p^d)^* \) such that \( A \cdot \| w \| = c \) only visit a finite number of states of \( A(S) \), and to obtain the \( p \)-automaton \( A(S) \) corresponding to \( S \) we subsequently restrict \( Q \) to only such states. The following lemma recalls an algebraic characterization of the reachability relation of \( A(S) \) established in the proof of Proposition 14 in [7].

**Lemma 1.** Let \( q, r \in \mathbb{Z}^m \) be states of \( A(S) \), \( w \in (\Sigma_p^d)^n \) and \( x = \| w \|_p \). Then \( q \xrightarrow{w} r \) if and only if there is \( y \in \mathbb{N} \) such that

\[
q = y \cdot r + A \cdot x, \quad \| x \|_\infty < y, \quad y = p^n.
\]

Let \( x \) be a distinguished variable of \( x \). For a word \( w \in (\Sigma_p^d)^n \) encoding solutions of \( S \), denote by \( \pi_x(w) \) the word \( v \in \Sigma_p^* \) obtained from projecting \( w \) onto the component of \( w \) corresponding to \( x \). Let \( q \) be a state of a \( p \)-automaton \( A \), define the counting function \( C_{q,x} : \mathbb{N} \to \mathbb{N} \) as

\[
C_{q,x}(n) := \# \{ \pi_x(w) : q \xrightarrow{w} q, w \in (\Sigma_p^d)^n \}.
\]

We now show that for \( p \)-automata arising from systems of linear Diophantine equations, \( C_{q,x} \) can be obtained from an eventual quasi-polynomial.

**Lemma 2.** For the \( p \)-automaton \( A(S) \) associated to \( S \): \( A \cdot x = c \) with states \( Q \) and all \( q \in Q \), there is an eventual quasi-polynomial \( f \) such that \( C_{q,x}(n) = f(p^n) \) for all \( n \in \mathbb{N} \). Moreover, for all sufficiently large \( n \in \mathbb{N} \), \( f_p^n \) is a linear polynomial.

**Proof.** Let \( q = q \in \mathbb{Z}^d \). By Lemma 1, \( q \xrightarrow{w} q \) for \( w \in (\Sigma_p^d)^n \) if and only if there is a \( y \in \mathbb{N} \) such that

\[
q = y \cdot q + A \cdot x, \quad \| x \|_\infty < y, \quad y = p^n,
\]

where \( x = \| w \|_p \). The set of solutions of \( S' : A \cdot x + y \cdot q = q, \| x \|_\infty < y \) is a hybrid linear set \( L(D, R) \subseteq \mathbb{N}^{d+1} \). Let \( L(B, P) \subseteq \mathbb{N}^2 \) be obtained from \( L(D, R) \) by projecting onto the components corresponding to \( x \) and \( y \), and assume that \( x \) corresponds to the first and \( y \) to the second component of \( L(B, P) \). Let \( M_t := \mathbb{N} \times \{ t \} \) and

\[
f(t) := \# (L(B, P) \cap M_t).
\]

Observe that \( C_{q,x}(n) = f(p^n) \) and that \( f(n) \) is finite for all \( n \in \mathbb{N} \) due to the constraint \( x < y \). Let \( P = \{ p_1, \ldots, p_k \} \), the following formula of parametric Presburger arithmetic defines \( L(B, P) \cap M_t \):

\[
\Phi_t(x, y) := \exists z_1 \cdots \exists z_k : \bigvee_{b \in B} \left( \frac{x}{b} = b + \sum_{i=1}^{k} z_i \cdot p_i \right) \land y = t
\]

Thus, \( f = \# \Phi_t(x, y) \) and, by application of Proposition 1, \( f \) is an eventual quasi-polynomial.

Since \( C_{q,x}(n) \leq p^n - 1 \) for all \( n \in \mathbb{N} \), we in particular have that all polynomials \( f_p^n \) constituting \( f \) are linear as they would otherwise outgrow \( C_{q,x} \). \( \square \)
The next step is to lift Lemma 2 to systems of linear Diophantine equations with valuation constraints. To this end, we define a DFA whose language encodes the set of all solutions of predicates of the form $V_p(x, y)$. Formally, for $S: V_p(x, y)$ we define $A(S) := (Q, \Sigma_p^d, \delta, q_0, F)$ such that

- $Q := \{0, 1\}$,
- $\delta(0, u) := 0$ for all $u \in \Sigma_p^d$ such that $\pi_x(u) = 0$,
- $\delta(1, u) := 1$ for all $u \in \Sigma_p^d$ such that $\pi_x(u) = 1$ and $\pi_y(u) > 0$,
- $q_0 := 0$, and
- $F := \{1\}$.

For $S: A \cdot x = c \land \bigwedge_{1 \leq i \leq \ell} V_p(x_i, y_i)$, we denote by $A(S)$ the DFA that can be obtained from the standard product construction on all DFA for the atomic properties:

**Lemma 3.** Let $S$ be a system of linear Diophantine equations with valuation constraints with the associated DFA $A(S)$ with states $Q$, and let $q \in Q$. There is an eventual quasi-polynomial $f$ such that $C_{q,x}(n) = f(p^n)$. Moreover, $f_{p^n}$ is a linear polynomial for all $n \in \mathbb{N}$.

**Proof.** Let $S: A \cdot x = c \land \bigwedge_{1 \leq i \leq \ell} V_p(x_i, y_i)$, we have $Q \subseteq \mathbb{Z}^m \times \{0, 1\}^\ell$ and thus $q = (q, b_1, \ldots, b_i) \in Q$. Any self-loop $q \xrightarrow{w} q$ with $q = (q, b_1, \ldots, b_i)$ is a self-loop for the DFA induced by the system of linear Diophantine equations $A \cdot x = c$ with the additional requirement that $\pi_x(\|w\|_p) = 0$ for all $1 \leq i \leq \ell$ and furthermore $\pi_y(\|w\|_p) = 0$ whenever $b_i = 1$. Thus $(q, 0) \xrightarrow{w} (q, 0)$ where

$$S': A \cdot x = c \land \bigwedge_{1 \leq i \leq \ell} x_i = 0 \land \bigwedge_{1 \leq i \leq \ell, b_i = 1} y_i = 0.$$ 

Conversely, $(q, 0) \xrightarrow{w} (q, 0)$ immediately gives $q \xrightarrow{w} q$. The statement is now an immediate consequence of the application of Lemma 2 to $S'$.

We will from now on implicitly apply Lemma 3. As a first application, we show that Lemma 3 allows us to classify the DFA associated to a system of linear Diophantine equations with valuation constraints.

**Lemma 4.** The DFA $A(S)$ associated to a system of linear Diophantine equations with valuation constraints $S$ with states $Q$ has either of the following properties:

(i) there is $q \in Q$ such that $C_{q,x}$ is an eventual quasi-polynomial $f$ and $f_{p^n}$ is a non-constant polynomial for infinitely many $n \in \mathbb{N}$; or

(ii) there is a constant $d \geq 0$ such that $C_{q,x}(n) \leq d$ for all $q \in Q$ and $n \in \mathbb{N}$.
Proof. Suppose $A(S)$ has Property (i). For a contradiction, suppose $d \geq 0$ exists. Let $f$ be the eventual quasi-polynomial from Property (i). Every non-constant polynomial $f_{p_n}$ constituting $f$ is of the form $a \cdot x + b$ with $a > 0$. As there are infinitely many such $n$, there is some linear polynomial $g(x) = a \cdot x + b$ such that $g = f_{p_n}$ for infinitely many $n \in \mathbb{N}$. Hence $g(p^n) > d$ for some sufficiently large $n \in \mathbb{N}$.

For the converse, suppose that $A(S)$ does not have Property (i). Then there are $\ell, m > 0$ such that all $f_{p^n}$ are constant polynomials bounded by some value $m \in \mathbb{N}$ for all $n \geq \ell$, $q \in Q$ and $f = C_{q,x}$. Hence we can choose $d = \max(\{C_{q,n}(n) : q \in Q, 0 < n \leq \ell\} \cup \{m\})$. \hfill $\square$

We are now in a position to prove a dichotomy of the growth of the number of words accepted by a DFA corresponding to a system of linear Diophantine equations with valuation constraints.

Lemma 5. Let $S$ be a fixed system of linear Diophantine equations with valuation constraints with the associated DFA $A(S)$. Let $L = \pi_x(L(A(S)))$, then either

(i) $d_L(n) \geq c \cdot p^n$ for some fixed constant $c > 0$ and infinitely many $n \in \mathbb{N}$; or
(ii) $d_L(n) = O(n^c)$ for some fixed constant $c \geq 0$.

Proof. Let $A(S)$ have the set of states $Q$, initial state $q_0$ and final state $q_f$. The DFA $A(S)$ has one of the two properties of Lemma 4.

If $A(S)$ has the property of Lemma 4(i) then consider $q \in Q$ such that $C_{q,x}$ is an eventual quasi-polynomial $f$ such that $f_{p^n}$ is non-constant for infinitely many $n \in \mathbb{N}$, and let $i_1 < i_2, \ldots, n$ be such that all $f_{p_{i_j}}$ are the same non-constant polynomial $a \cdot x + b$. Consider $v$ and $w$ such that $q_0 \xrightarrow{v} q \xrightarrow{w} q_f$. Then for all sufficiently large $j$ we have

$$d_L(i_j + |v| + |w|) \geq a \cdot p^{i_j} + b \geq c \cdot p^{(i_j + |v| + |w|)}$$

for some fixed constant $c > 0$.

Otherwise, $A(S)$ has the property of Lemma 4(ii), and there is some fixed $d \geq 0$ such that $C_{q,x}(n) \leq d$ for all $n \in \mathbb{N}$ and $q \in Q$. Every $w \in L$ such that $|w| = n$ can uniquely be decomposed as $w = v_0 w_1 v_1 w_2 \cdots w_k v_k$ for some $k \leq |Q|$ such that

$$q_0 \xrightarrow{v_0} q_{a_1} \xrightarrow{w_1} q_{a_2} \xrightarrow{v_1} q_{a_2} \xrightarrow{w_2} q_{a_2} \xrightarrow{v_2} q_{a_3} \cdots \xrightarrow{w_k} q_{a_k} \xrightarrow{v_k} q_{a_{k+1}}.$$  \hfill (1)

where $q_{a_{k+1}} = q_f$, $q_{a_i} \neq q_{a_j}$ for all $i \neq j$ and each $q_{a_i} \xrightarrow{v_i} q_{a_{i+1}}$ corresponds to a loop-free path in $A(S)$. Since $C_{q,x} \leq d$, there are at most $d^k \leq d^{\#Q}$ words $u \in L$ of length $n$ that have the same sequence of states in the decomposition of Eq. (1) at the same position where they occur in $w$. Moreover, there are at most $\binom{n}{2k} \leq n^{\binom{\#Q}{2}} \leq n^{(n+\#Q)}$ possibilities at which the states $q_{a_i}$ can appear in any $u \in L$ of length $n$ for any particular sequence of states in the decomposition of Eq. (1). Finally, there are at most $n^{(\#Q)\#Q}$ such sequences. We thus derive:

$$d_L(n) \leq (\#Q)^\#Q \cdot n^{(2\#Q)} \cdot d^{\#Q} = O(n^c)$$

for some constant $c \geq 0$. \hfill $\square$
Corollary 1. Let $\Phi(x)$ be a fixed formula of existential Büchi arithmetic of base $p$. Let $M = \llbracket \Phi(x) \rrbracket_p$, then either:

(i) $d_M(n) \geq c \cdot p^n$ for some fixed constant $c > 0$ and infinitely many $n \in \mathbb{N}$; or

(ii) $d_M(n) = O(n^c)$ for some fixed constant $c \geq 0$.

Proof. Without loss of generality we may assume that $\Phi(x)$ is in disjunctive normal form such that $\Phi(x) = \bigvee_{i \in I} \Phi_i(x)$ and each $\Phi_i(x)$ is a system of linear Diophantine equations with valuation constraints $S_i$. For $M_i = \llbracket \Phi_i(x) \rrbracket_p$, we obtain $d_{M_i}$ by application of Lemma 5. If there is a constant $c \geq 0$ such that $d_{M_i} = O(n^c)$ for all $i \in I$ then $d_M = O(n^c)$. Otherwise, if there is some $i \in I$ such that $d_{M_i}(n) \geq c \cdot p^n$ for some constant $c > 0$ and infinitely many $n \in \mathbb{N}$ then $d_M(n) \geq c \cdot p^n$ for infinitely many $n \in \mathbb{N}$. □

As an immediate consequence of Corollary 1, we obtain:

Corollary 2. Let $M \subseteq \mathbb{N}$ such that $f = o(d_M)$ for any $f = O(n^c)$, $c \geq 0$, and $d_M = o(p^n)$. Then $M \not\subseteq \Sigma_1$-$\text{BA}$.

For any $p \geq 2$, consider $L = (10 | 01)^* \subseteq \Sigma_p^*$ and $M = \llbracket L \rrbracket_p$. We have $d_M(n) \leq 2^{n/2}$ and

$$\lim_{n \to \infty} \frac{f(n)}{d_M(n)} = 0 \quad \lim_{n \to \infty} \frac{d_M(n)}{c \cdot p^n} = 0$$

for any $f = O(n^c)$ and $c \geq 0$. On the one hand, Corollary 2 yields that $M \not\subseteq \Sigma_1$-$\text{BA}$. On the other hand, since $M$ is $p$-regular, we have $M \in \text{BA}$. This concludes the proof of Theorem 1.

4 Expressive completeness of the $\Sigma_2$-fragment of Büchi arithmetic

For a regular language $L \subseteq (\Sigma_p^d)^*$ given by a DFA, Villemaire shows in the proof of Theorem 2.2 in [12] how to construct a $\Sigma_3$-formula of Büchi arithmetic $\Phi_L(x)$ such that $\llbracket \Phi_L(x) \rrbracket_p = \llbracket L \rrbracket_p$. This construction is modularized and relies on an existential formula $\Phi_{p,j}(x,y)$ expressing that “$x$ is a power of $p$ and the coefficient of this power of $p$ in the representation of $y$ in base $p$ is $j$”:

$$\Phi_{p,j}(x,y) \equiv P_p(x) \land \exists x \exists u \exists z: (y = z + j \cdot y + t) \land (z < x) \land (V_p(u, t) \land x < u \lor t = 0).$$

The only reason why $\Phi_L(x)$ in [12] is a $\Sigma_3$-formula is that $\Phi_{p,j}(x,y)$ appears in an implication both as antecedent and as consequent inside an existential formula. Thus, if one could additionally define $\Phi_{p,j}(x,y)$ by a $\Pi_1$-formula then $\Phi_L(x)$ immediately becomes a $\Sigma_2$-formula. That is, however, not difficult to achieve by defining:

$$\tilde{\Phi}_{p,j}(x,y) \equiv P_p(x) \land \forall s \forall t \forall u \forall v:$$

$$\left( \neg(s = z + j \cdot x + t) \lor (z \geq x) \lor (V_p(u, t) \lor x \geq u) \land \neg(t = 0) \right) \rightarrow \neg(s = y).$$
Note that the order relation can also be expressed by a universal formula: $x \leq y$ if and only if $\forall z: (y + z = x) \rightarrow (z = 0)$. Thus, $\Phi_{p,j}(x,y)$ is indeed a $\Pi_2$ formula.

Combining $\tilde{\Phi}_{p,j}(x,y)$ with the results in [12], we obtain that the $\Sigma_2$-fragment of Büchi arithmetic is expressively complete.

**Theorem 2.** For any base $p$, $\Sigma_2$-$BA_p = BA_p$.

5 **Existential Büchi arithmetic defines regular languages of polynomial growth**

For a language $L \subseteq \Sigma^*$, Szilard et al. [10] say that $L$ has polynomial growth if $d_L(n) = O(n^c)$ for some constant $c \geq 0$ and all $n \in \mathbb{N}$. Denote by

$$\text{PREG}_p := \{[L]_p : L \subseteq \Sigma^p, \text{ $L$ is a regular language of polynomial growth}\}$$

the numerical encoding of all regular languages of polynomial growth in base $p$. One of the main results of [10] is that a regular language $L$ has polynomial growth if and only if $L$ can be represented as a finite union of regular expressions of the form

$$v_0 w_1^* v_1 \cdots v_{k-1} w_k^* v_k.$$  (2)

We show in this section that existential Büchi arithmetic defines any regular language of the form in Eq. (2). This immediately gives the following theorem.

**Theorem 3.** For any base $p$, $\text{PREG}_p \subseteq \Sigma_1$-$BA_p$.

We first require a couple of abbreviations. Define

$$W_p(x,y) \equiv P_p(y) \land x < y \leq p \cdot x,$$

which expresses that $y$ is the smallest power of $p$ strictly greater than $x$.

Let $\ell > 0$, Lohrey and Zetzsche introduce in [8] the predicate $S_\ell(x,y)$ which holds whenever

$$x = p^r \text{ and } y = p^{r+\ell \cdot i} \text{ for some } i, r \geq 0.$$

They show that $S_\ell(x,y)$ is definable in existential Büchi arithmetic. Since $y = p^{r+\ell \cdot i} \cdot x$ if and only if $y \equiv x \mod (p^\ell - 1)$, one can obtain $S_\ell$ as

$$S_\ell(x,y) \equiv P_p(x) \land P_p(y) \land \exists z: (y - x = (p^\ell - 1) \cdot z) \land y \geq x.$$

We slightly generalize $S_\ell$. Let $U \subseteq \mathbb{N}$, define the predicate $S_U(x,y)$ to hold whenever

$$x = p^r \text{ and } y = p^{r+u} \text{ for some } r \geq 0 \text{ and } u \in U.$$

**Lemma 6.** For any ultimately periodic set $U \subseteq \mathbb{N}$, the predicate $S_U(x,y)$ is definable in existential Büchi arithmetic.
Proof. Suppose that $U$ is given as $(t, \ell, B, R)$, we define $S_U(x, y) \equiv P_p(x) \land P_p(y) \land \bigvee_{b \in B} y = p^b \cdot x \lor \bigvee_{r \in R} S_t(p^{t+r} \cdot x, y)$.

\[\square\]

Towards proving Theorem 3, we now show that we can define $J^w \cdot K_p$ for any $w \in \Sigma_p$.

Lemma 7. For any $w \in \Sigma_p^*$, $[w^*]_p$ is definable by a formula of existential Büchi arithmetic $\Phi_{w^*}(x)$.

Proof. Let $m = p^r$ be the smallest power of $p$ greater than $[w]_p$. Then for any $k \geq 0$, $[w^k]_p = [w]_p \cdot \sum_{i=0}^{k-1} m^i = [w]_p \cdot \frac{m^k - 1}{m - 1}$.

It follows that $[w^*]_p$ is defined by $\Phi_{w^*}(x) \equiv \exists y \exists z : y < z \land \Phi_{w^*}(y) \land \Phi_{w^*}(z) \land \bigvee_{0 \leq i < |w|} x = p^i \cdot (z - y) \land \exists s \exists t : S_U(1, s) \land V_p(t, x) \land t = p^{[w_0]_p + 1} \cdot s$.

\[\square\]

Building upon Lemma 7, we now show that, for any $w \in \Sigma_p$, we can define $[w^+]_p$ shifted to the left by a number of zeros specified by an ultimately periodic set.

Lemma 8. Let $w \in \Sigma_p^*$ and $U$ be an ultimately periodic set. Then $[w^+U]_p$ is definable by a formula of existential Büchi arithmetic $\Phi_{U,w^+}(x)$.

Proof. The case $w \in 0^*$ is trivial. Thus, let $w = w' \cdot w_0$ such that $w' \in \Sigma_p^* \setminus \{0\}$ and $w_0 \in 0^*$. Observe that for $i < j$, $[w^i]_p - [w^j]_p = [w^{i-j}0^j]_p$. We define $\Phi_{U,w^+}(x) \equiv \exists y : S_U(1, y) \land x = [w]_p \cdot (y - 1)$.

\[\square\]

We have now all the ingredients to prove the following key proposition. In the proof below, recall that $W_p(x, y)$ holds if and only if $y$ is the smallest power of $p$ strictly greater than $x$.

Proposition 2. Let $L = v_0 w^*_1 v_1 \cdots v_{k-1} w^*_k v_k$. Then $[L]_p$ is definable in existential Büchi arithmetic.
Proof. The proposition follows from showing the statement for languages of the form
\[ L' = v_0 w_1^* v_1 \cdots v_{k-1} w_k^* v_k. \]

We show the statement by induction on \( k \). The induction base case \( k = 0 \) is trivial. For the induction step, assume that for \( M = v_1 w_2^* v_2 \cdots v_{k-1} w_k^* v_k \), \([M]_p \) is defined by a formula \( \Phi_k(x) \) of existential Büchi arithmetic, and let \( v_0, w_1 \in \Sigma_p^* \).

We first show how to define \( N = w_1^* v_1 w_2^* v_2 \cdots v_{k-1} w_k^* v_k \). To this end, factor \( M = M_0 \cdot M' \), where \( M_0 \subseteq 0^* \) and \( M \subseteq (\Sigma_p \setminus \{0\}) \cdot \Sigma_p^* \). Observe that \([M]_p = [\Phi_k(x)]_p \) and that both \( U = \{ |w| : w \in M \} \) and \( V = \{ |w| : w \in M_0 \} \) are ultimately periodic sets, cf. \([5,11]\). We moreover assume that \( w_1 \not\in 0^* \), otherwise we are done. Factor \( w_1 = w' \cdot w_0 \) such that \( w' \in \Sigma_p^* \cdot (\Sigma_p \setminus \{0\}) \) and \( w_0 \in 0^* \).

Now define
\[
\Psi_{k+1}(x) \equiv \exists y \exists z : \Phi_k(y) \land \Phi_{U',w^+}(z) \land x = y + z ^\land \\
\land \exists s \exists t : W_p(y, s) \land S_V(s, t) \land V_p(p^{|w_0|+1}, t, z). 
\]

The first line composes \( x \) as the sum of some \( y \in [M]_p \) and \( z \in [w^+0^J]_p \). The next line ensures that the number of zeros between the leading bit of \( y \) and last non-zero digit of \( z \) in their \( p \)-ary expansion is in \( V + |w_0| \). Thus, \([N]_p = [\Psi_{k+1}(x)]_p \).

We now show how to define \( L' \) along similar lines. To this end, factor \( N = N_0 \cdot N' \) such that \( N_0 \subseteq 0^* \) and \( N' \subseteq (\Sigma_p \setminus \{0\}) \cdot \Sigma_p^* \), and let \( T = \{ |w| : w \in N_0 \} \), which is an ultimately periodic set. We now obtain the desired formula of existential Büchi arithmetic as
\[
\Phi_{k+1}(x) \equiv \exists y \exists z : x = y + p \cdot z \cdot [v_0]_p \land \Psi_{k+1}(y) \land \exists s : W_p(y, s) \land S_T(s, z). 
\]

Since we can define any regular language of the form \((2)\) in existential Büchi arithmetic via Proposition 2, we can define a finite union of such languages and thus define all regular languages of polynomial growth in existential Büchi arithmetic. This completes the proof of Theorem 3.

Note that \( \text{PREG}_p \not\subseteq \text{PA} \) for any base \( p \geq 2 \); since \( M = [\Phi(x)] \) is ultimately periodic for any formula \( \Phi(x) \) of Presburger arithmetic, it follows that \( d_M(n) = \Omega(p^n) \), i.e., not of polynomial growth.

6 Conclusion

The main result of this paper is that existential Büchi arithmetic is strictly less expressive than full Büchi arithmetic of any base. This in contrast to Presburger arithmetic, for which it is known that its existential fragment is expressively complete.

When considered as the first-order theory of the structure \( (\mathbb{N}, 0, 1, +) \), Presburger arithmetic does not have a quantifier elimination procedure. The extended
structure \(\langle \mathbb{N}, 0, 1, +, \{c \cdot \} \rangle_{c > 1}\), on the other hand, admits quantifier elimination. Those additional divisibility predicates are definable in existential Presburger arithmetic. Our main result shows that even if we extended the structure underlying Büchi arithmetic with predicates definable in existential Büchi arithmetic, the resulting first-order theory would not admit quantifier-elimination. It would be interesting to know whether there is a non-trivial extension of Büchi arithmetic that does not increase the expressive power yet allows for a quantifier-elimination procedure.

Acknowledgments. We would like to thank Dmitry Chistikov and Alex Fung for inspiring discussions on the topics of this paper.

References

1. V. Bruyère, *Entiers et automates finis*, Mémoire de fin d’études (1985).
2. V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire, Logic and \(p\)-recognizable sets of integers, Bull. Belg. Math. Soc. Simon Stevin 1 (1994), no. 2, 191–238.
3. J.R. Büchi, Weak second-order arithmetic and finite automata, Math. Logic Quart. 6 (1960), no. 1-6, 66–92.
4. D. Chistikov and C. Haase, The taming of the semi-linear set, Automata, Languages, and Programming, ICALP, LIPIcs, vol. 55, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016, pp. 128:1–128:13.
5. M. Chrobak, Finite automata and unary languages, Theor. Comput. Sci. 47 (1986), no. 3, 149–158.
6. S. Ginsburg and E.H. Spanier, Bounded ALGOL-like languages, T. Am. Math. Soc. (1964), 333–368.
7. F. Guépin, C. Haase, and J. Worrell, On the existential theories of Büchi arithmetic and linear \(p\)-adic fields, Logic in Computer Science, LICS, IEEE, 2019, pp. 1–10.
8. M. Lohrey and G. Zetzsche, Knapsack and the power word problem in solvable Baumslag-Solitar groups, Mathematical Foundations of Computer Science, MFCS, LIPIcs, vol. 170, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 67:1–67:15.
9. M. Presburger, Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt, Comptes Rendus du I congres de Mathematiciens des Pays Slaves, 1929, pp. 92–101.
10. A. Szilard, S. Yu, K. Zhang, and J.O. Shallit, Characterizing regular languages with polynomial densities, Mathematical Foundations of Computer Science, MFCS, Lect. Notes Comp. Sci., vol. 629, Springer, 1992, pp. 494–503.
11. A.W. To, Unary finite automata vs. arithmetic progressions, Inf. Process. Lett. 109 (2009), no. 17, 1010–1014.
12. R. Villemaire, The theory of \(\langle \mathbb{N}, +, V_k, V_l \rangle\) is undecidable, Theor. Comput. Sci. 106 (1992), no. 2, 337–349.
13. P. Wolper and B. Boigelot, On the construction of automata from linear arithmetic constraints, Tools and Algorithms for the Construction and Analysis of Systems, TACAS, Lect. Notes Comp. Sci., vol. 1785, Springer, 2000, pp. 1–19.
14. K. Woods, The unreasonable ubiquitouenss of quasi-polynomials, Elect. J. Combin. 21 (2014), no. 1, P1.44.