Canad. J. Math. 2026, pp. 1–27
http://dx.doi.org/10.4153/S0008414X2610234X
© The Author(s), 2026. Published by Cambridge University Press on behalf of Canadian Mathematical Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

# A dichotomy for $k$-automatic expansions of Presburger arithmetic

Jason Bell, Alexi Block Gorman, and Chris Schulz

*Abstract.* Let $k \geq 2$ and let $X$ be a subset of the natural numbers that is $k$-automatic and not eventually periodic. We show that the following dichotomy holds: either all $k$-automatic subsets are definable in the expansion of Presburger arithmetic in which we adjoin the predicate $X$, or $(\mathbb{N}, +, X)$ has the same definable sets as $(\mathbb{N}, +, k^{\mathbb{N}})$.

## 1 Introduction

A set $X \subseteq \mathbb{N}$ is *$k$-automatic* if the language of base-$k$ expansions of elements of $X$, viewed as strings over the alphabet $\{0, 1, \ldots, k-1\}$, is accepted by some finite-state automaton. These sets, and their definability properties, have played an important role in both model theory and in theoretical computer science. A key structural insight, due to Büchi in [10] and Bruyère in [8], is that the $k$-automatic subsets of $\mathbb{N}^n$ are exactly those definable in the structure $(\mathbb{N}, +, V_k)$, where the unary function $V_k : \mathbb{N} \to \mathbb{N}$ returns the largest power of $k$ that divides $n$.

This correspondence has far-reaching consequences: it shows, for example, that $(\mathbb{N}, +, V_k)$ admits a decidable theory. The broader project of understanding expansions of Presburger arithmetic by automatic sets has a long and rich history. One such result is the Cobham–Semenov theorem [11] which asserts that if a set $X \subseteq \mathbb{N}^n$ is definable both in $(\mathbb{N}, +, V_k)$ and in $(\mathbb{N}, +, V_\ell)$, for multiplicatively independent bases $k$ and $\ell$, then $X$ is already definable in Presburger arithmetic; that is, it is definable in $(\mathbb{N}, +)$. Other such results include the fact that the structure $(\mathbb{N}, +, V_k, V_\ell)$ defines multiplication when $k$ and $\ell$ are multiplicatively independent, due to Villemaire in [19], which was then strengthened by Bès in [7], in which he proves that $(\mathbb{N}, +, V_k, \ell^{\mathbb{N}})$ defines multiplication when $k$ and $\ell$ are multiplicatively independent.

There exists a natural hierarchy of structures expanding Presburger arithmetic by non-Presburger-definable $k$-automatic sets. At the top lies $(\mathbb{N}, +, V_k)$, which completely captures $k$-automaticity and exhibits some degree of model-theoretic complexity (e.g., it has TP $_2$). At the bottom lies $(\mathbb{N}, +, k^{\mathbb{N}})$, where $k^{\mathbb{N}} = \{k^n : n \in \mathbb{N}\}$, a reduct that has NIP and in which the definable sets are comparatively tame (see Lambotte–Point [15, Theorem 2.32] and Semenov [18]).

Received by the editors August 21, 2025 UTC; revised May 5, 2026 UTC; accepted June 8, 2026 UTC.
Published online on Cambridge Core June 24, 2026.
The first-named author was supported by NSERC grant RGPIN-2022-02951. The second-named author was supported by the National Science Foundation under Award No. DMS-2303368.
AMS subject classification: **03C64**, **03D05**.
Keywords: Presburger, finite automata, model theory, definability, decidability.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

In this article, we show that for expansions of $(\mathbb{N}, +)$ by a single unary $k$-automatic set $X \subseteq \mathbb{N}$ that is not eventually periodic, the hierarchy collapses. Specifically, we prove the following dichotomy holds.

***Theorem 1.1*** *Let $X \subseteq \mathbb{N}$ be a $k$-automatic set that is not eventually periodic. Then either $(\mathbb{N}, +, X)$ defines all $k$-automatic sets, or $X$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$.*

The main ingredients for proving this theorem are characterizations of sparse and non-sparse languages, as well as the definable sets of $(\mathbb{N}, +, k^{\mathbb{N}})$ and $(\mathbb{N}, +)$, respectively. In particular, we use that sets definable from sparse ones are precisely those definable in $(\mathbb{N}, +, k^{\mathbb{N}})$. On the other hand, sets not definable from a sparse one necessarily define an additive basis[1] for the natural numbers that is not eventually periodic and from such a set we show that one can define all automatic sets. We also make key use of the fact that to check whether a $k$-automatic set $X$ is Presburger-definable (i.e., definable in $(\mathbb{N}, +)$) it suffices to check only that it has period $p \geq 1$ on an interval that is bounded in terms of $p$ and $X$.

The outline of this article is as follows. In Section 2, we provide background on automata and Presburger definability, in addition to other relevant concepts required to prove Theorem 1.1. In Section 3, we prove Theorem 1.1 in a key special case: namely, when $X$ is the set of natural numbers whose base-$k$ expansions are accepted by a cycle language; that is, a regular language accepted by a deterministic finite-state automaton with one final state that is equal to its initial state (see Section 2 for definitions). In Section 4, we prove the general case. Finally, in Section 5, we make some observations that follow from our main result and a direction for further inquiry.

## 2 Preliminaries

### 2.1 Presburger arithmetic

Presburger arithmetic refers to the first-order theory of natural numbers with addition. It was introduced by Presburger in [16], who provided a recursive axiomatization and proved that the theory is both consistent and complete.

***Definition 2.1*** *Presburger arithmetic* is the first-order theory of the structure $(\mathbb{N}, +)$.

Throughout this article, we consider first-order formulas in Presburger arithmetic only in terms of the sets they define in the standard model. Non-standard models will not concern us.

There is a number-theoretic characterization of the sets definable in Presburger arithmetic, due to Ginsburg and Spanier [13].

***Definition 2.2*** A subset $X$ of $\mathbb{N}^d$ is said to be *linear* if it is of the form:

$$X = \left\{ x + \sum_{i=1}^{n} k_i y_i : k_1, \ldots, k_n \in \mathbb{N} \right\},$$

where $x, y_1, \ldots, y_n$ are fixed vectors in $\mathbb{N}^d$.

[1] A set $X$ such that for some $n \in \mathbb{N}$, every number can be written as a sum of at most $n$ elements of $X$.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

A subset of $\mathbb{N}^d$ is *semilinear* if it is a finite union of linear sets.

***Theorem 2.3*** [13] *A subset of $\mathbb{N}^d$ is definable in Presburger arithmetic if and only if it is semilinear.*

In this article, we focus on the one-dimensional case. In this setting, semilinear sets correspond precisely to eventually periodic sets.

***Definition 2.4*** A subset $X$ of $\mathbb{N}$ is *eventually periodic* if there exist $N, p \in \mathbb{N}$ such that $n \in X$ if and only if $n + p \in X$ for all $n \geq N$.

Thus, we have the following result, which also follows from quantifier elimination for Presburger arithmetic (see Theorem 32F in [12]).

***Corollary 2.5*** *Let $X \subseteq \mathbb{N}$. Then $X$ is Presburger-definable if and only if $X$ is eventually periodic.*

## 2.2 Finite automata

This article will study regular languages by way of the corresponding finite automata. As such, we begin with a definition of the non-deterministic finite automaton (NFA).

***Definition 2.6*** An NFA is a tuple $\mathcal{A} = (Q, I, \Sigma, \delta, F)$ where:

- $Q$ is a finite set of elements called *states*;
- $I \subseteq Q$ is the *set of start states* or *initial states*;
- $\Sigma$ is a finite *alphabet*;
- $\delta : Q \times \Sigma \to \mathcal{P}(Q)$ is a function called the *transition function*;
- $F \subseteq Q$ is a set of *accept states* or *final states*.

***Definition 2.7*** An NFA $\mathcal{A} = (Q, I, \Sigma, \delta, F)$ is called *deterministic*, or a *DFA*, if $I$ has exactly one element and if $\delta(q, y)$ has at most one element for all $q \in Q, y \in \Sigma$.

We first recall that for an NFA $\mathcal{A} = (Q, I, \Sigma, \delta, F)$, we can inductively extend the function $\delta : Q \times \Sigma \to \mathcal{P}(Q)$ to a function from $Q \times \Sigma^*$ to $\mathcal{P}(Q)$ by declaring that, for a word $v \in \Sigma^*$ and $y \in \Sigma$:

$$\delta(q, vy) = \bigcup_{q' \in \delta(q,v)} \delta(q', y).$$

If $\mathcal{A}$ is deterministic, this extended $\delta$ maintains the same property that its output is always at most a singleton set.

A *run* of $\mathcal{A}$ on a word $w \in \Sigma^*$ is a finite sequence $(q_0, q_1, \ldots, q_{|w|})$ of states, with $q_0 \in I$ a start state, such that, for $1 \leq i \leq |w|$, we have $q_i \in \delta(q_{i-1}, w_i)$. A run is *accepting* if $q_{|w|} \in F$; if any accepting run exists, we say that $\mathcal{A}$ *accepts* $w$. If $\mathcal{A}$ does not accept a word $w$, we say that it *rejects* $w$. We may also speak of runs *from* states other than initial states; a run from $q$ means a run in the automaton $\mathcal{A}'$ that replaces $I$ with $\{q\}$.

A language $L$ is regular if and only if there exists a finite automaton $\mathcal{A}$ such that $L$ is the set of words accepted by $\mathcal{A}$. It is a well-known fact due to Rabin and Scott that we may without loss of generality take $\mathcal{A}$ to be deterministic.

***Fact 2.8*** *Let $\mathcal{A}$ be any NFA. There exists a deterministic finite automaton $\mathcal{A}'$ such that $\mathcal{A}$ and $\mathcal{A}'$ accept precisely the same words.*

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

We will also need two facts about the sizes of deterministic finite automata after transformations are applied to the corresponding languages.

***Lemma 2.9*** *Let $L$ be a regular language whose deterministic finite automaton $\mathcal{A}$ has $M$ states. Let $u$ be any word. Then the language $\{v : uv \in L\}$ has a deterministic finite automaton with $(M+1)$ states.*

**Proof** Let $\mathcal{A} = (Q, I, \Sigma, \delta, F)$. We define the DFA $\mathcal{A}'$ on $\Sigma$ as follows. First, add a new state $q_u$ to $Q$ to obtain the set of states in $\mathcal{A}'$; $q_u$ will be the initial state. The final states of $\mathcal{A}'$ will be those in $F$; if (and only if) $u \in L$, we also add $q_u$ as a final state. Finally, the transition function of $\mathcal{A}'$ agrees with that of $\mathcal{A}$ on its domain, and for each symbol $\sigma \in \Sigma$, we add a transition from $q_u$ to $\delta(I, u\sigma)$ on $\sigma$.

We claim that $\mathcal{A}'$ recognizes $\{v : uv \in L\}$. Let $uv \in L$. If $v$ is empty, then $u \in L$ and hence the empty word is recognized by $\mathcal{A}'$. Otherwise, let $v = \sigma w$. Then $\delta(I, u\sigma w) \in F$; so $\delta(\delta(I, u\sigma), w) \in F$; so $\delta(q_u, \sigma w) \in F$, as needed. The proof when $uv \notin L$ is analogous. ∎

***Lemma 2.10*** *Let $L_1$ be a regular language whose deterministic finite automaton $\mathcal{A}_1$ has $M_1$ states, and let $L_2$ be a regular language whose deterministic finite automaton $\mathcal{A}_2$ has $M_2$ states. Then the language $\{w : w \in L_1 \leftrightarrow w \in L_2\}$ has a deterministic finite automaton with $M_1 \cdot M_2$ states.*

**Proof** Let $\mathcal{A}_1 = (Q_1, I_1, \Sigma, \delta_1, F_1)$ and $\mathcal{A}_2 = (Q_2, I_2, \Sigma, \delta_2, F_2)$. We will construct a new automaton $\mathcal{A}_3 = (Q_3, I_3, \Sigma, \delta_3, F_3)$. Let $Q_3 = Q_1 \times Q_2$, $I_3 = I_1 \times I_2$, and let $\delta_3((q_1, q_2), \sigma) = \delta_1(q_1, \sigma) \times \delta_2(q_2, \sigma)$. Note that this definition naturally extends to words; $\delta_3((q_1, q_2), w) = \delta_1(q_1, w) \times \delta_2(q_2, w)$.

Let $F_3 = \{(q_1, q_2) \in Q_3 : q_1 \in F_1 \leftrightarrow q_2 \in F_2\}$. Then we note that $\mathcal{A}_3$ recognizes $w$ iff the single element of $\delta_3((q_1, q_2), w)$ is in $F_3$. This is true if and only if the equivalence $\delta_1(q_1, w) \in F_1 \leftrightarrow \delta_2(q_2, w) \in F_2$ holds; but this is precisely when $w \in L_1 \leftrightarrow w \in L_2$. ∎

Call an NFA *trim* if every state can be reached by some path from the start state and, for each state, there exists a final state that can be reached from the state in question. Call an NFA or DFA recognizing the language $L$ *minimized* if it has the smallest number of states possible among all automata that recognize $L$ (within the class of NFAs and DFAs, respectively). Note that for any NFA, there exists an equivalent trim automaton and an equivalent minimized automaton, and these can be made deterministic up to allowing the transition function to be a partial function.

***Definition 2.11*** The *digraph structure* of a finite automaton $\mathcal{A} = (Q, I, \Sigma, \delta, F)$ is the directed graph $(Q, \Delta)$, where $p \Delta q \iff \exists \sigma \in \Sigma : q \in \delta(p, \sigma)$.

We may use digraph terminology to apply properties of the digraph structure to the automaton. For example, a *strongly connected component* of $\mathcal{A}$ is a maximal subset $S \subseteq Q$ such that, for all $p, q \in S$, there exists a path from $p$ to $q$, or equivalently a word $w$ such that $\mathcal{A}$ runs from $p$ to $q$ on $w$. We define a *leaf* of the digraph structure to be a strongly connected component with no outgoing transitions to states in a different strongly connected component.

Within this article, it will be very useful to examine the internal structure of an automaton in terms of which words run to which individual states. So we will define two classes of languages.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

***Definition 2.12*** Let $\mathcal{A} = (Q, I, \Sigma, \delta, F)$ be a finite automaton and $p, q \in Q$. The *path language* $L_{p\to q} \subseteq \Sigma^*$ is the set of words $w$ on which there is a run from $p$ to $q$ in $\mathcal{A}$. In the case that $p = q$, we call this the *cycle language*.

It is elementary that path languages are always regular. Moreover, note that $L_{q\to q} = L_{q\to q}^*$.

***Definition 2.13*** An *induced subautomaton* of an automaton $\mathcal{A} = (Q, I, \Sigma, \delta, F)$ corresponding to some $Q' \subseteq Q$ is the automaton $(Q', I', \Sigma, \delta|_{Q'}, F \cap Q')$, where $I'$ is the union of $Q' \cap I$ along with all elements of $Q'$ to which there is an incoming transition from outside of $Q'$.

## 2.3 Encoding sets of natural numbers

Our results in this article relate to the encoding of natural numbers via the base-$k$ expansion. We define this, and our notations related to it, here.

***Definition 2.14*** Given a natural number $k \geq 2$, we let $\Sigma_k = \{0, \ldots, k-1\}$. A *base-k expansion* of a natural number $x$ is a word $w_n w_{n-1} \ldots w_0$ over the alphabet $\Sigma_k$ such that:

$$x = \sum_{i=0}^{n} w_i k^i.$$

In this case, we say that $x = [w]_k$.

Note that the above definition defines $[w]_k$ by treating $w$ as a *most-significant-digit-first* (MSD-first) representation. Some previous work has used the opposite *least-significant-digit-first* representation, so in order to ensure that the reader understands what our convention is in this article, we instruct the reader to assume that MSD-first representations are in use in all cases. By example: $[120]_3 = 1 \cdot 3^2 + 2 \cdot 3^1 = 15$, not 7.

We will commonly apply this notation to sets as well. So given a language $L \subseteq \Sigma_k^*$, we let $[L]_k$ denote the set of natural numbers of the form $[w]_k$ with $w \in L$. We say that a set $X \subseteq \mathbb{N}$ is *k-automatic* if there is a regular sublanguage $L$ of $\Sigma_k^*$ such that $X = [L]_k$.

We note that there is a well-known dichotomy for regular languages: if $L$ is a regular language and if $|L_{\leq n}|$ denotes the number of words in $L$ of length at most $n$, then either $|L_{\leq n}|$ is polynomially bounded in $n$, or else there is a constant $C > 1$ such that $|L_{\leq n}| \geq C^n$ for infinitely many $n$. More generally, this dichotomy can be stated as follows.

***Theorem 2.15*** [6, Proposition 7.1] *Let $L$ be a regular language. The following are equivalent:*

(1) *$f_{\mathcal{L}}(n) = O(n^d)$ for some natural number $d$.*
(2) *$f_{\mathcal{L}}(n) = o(C^n)$ for every $C > 1$.*
(3) *There do not exist words $u, v, a, b$ with $a, b$ nontrivial and of the same length and $a \neq b$ such that $u\{a, b\}^* v \subseteq L$.*
(4) *Suppose $\Gamma = (Q, \Sigma, \delta, q_0, F)$ is an automaton accepting $L$ in which all states are accessible. Then $\Gamma$ satisfies the following:*

   *$(*)$ If $q$ is a state such that $\delta(q, v) \in F$ for some word $v$, then there is at most one nontrivial word $w$ with the property that $\delta(q, w) = q$ and $\delta(q, w') \neq q$ for every nontrivial proper prefix $w'$ of $w$.*

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

(5) *There exists an automaton accepting L that satisfies* $(*)$.
(6) *L is a finite union of languages of the form* $v_1 w_1^* v_2 w_2^* \dots v_m w_m^* v_{m+1}$, *where* $k \geq 0$ *and the* $v_i$*'s are possibly trivial words and the* $w_i$*'s are nontrivial words.*

We say that a regular language $L$ is *sparse* if $|L_{\leq n}| \in O(n^d)$ for some natural number $d$, and we will say that a $k$- automatic subset $X$ of $\mathbb{N}$ is *sparse* if $X = [L]_k$ for some sparse regular sublanguage of $\Sigma_k^*$.

There is a fundamental connection between $k$-automatic sets and the definable sets in an expansion of Presburger arithmetic, which was discovered by Büchi [9] and formally proven by Bruyère [8].

***Fact 2.16*** *Let* $V_k$ *send every positive integer* $x$ *to the largest power of* $k$ *dividing* $x$. *A set* $X \subseteq \mathbb{N}$ *is* $k$*-automatic if and only if it is definable in the structure* $(\mathbb{N}, +, V_k)$.

Therefore, the definable sets in $(\mathbb{N}, +, V_k)$ can be understood via the theory of finite automata. This result is fundamental to the study of base-$k$ arithmetic; for instance, it gives us an effective correspondence between formulas in the language of $(\mathbb{N}, +, V_k)$ and finite automata, and hence a decision algorithm in this language. The reducts of this structure are the subject of this article.

We shall need one more result, proven by Bès in [7].

***Fact 2.17*** [7, Theorem 3.1] *Let* $X \subseteq \mathbb{N}$ *be* $k$*-automatic but not definable in* $(\mathbb{N}, +)$ *(i.e., not eventually periodic). Then the structure* $(\mathbb{N}, +, X)$ *defines the set* $k^{\mathbb{N}} = \{k^n : n \in \mathbb{N}\}$.

## 2.4 Semenov's characterization

Semenov [18] gives a characterization of the one-dimensional sets definable from $(\mathbb{N}, +, k^{\mathbb{N}})$, which will be very useful to us. We first define the following subclass of regular languages.

***Definition 2.18*** Let $k$ be a natural number. A *$k$-Semenov-type* language is a sublanguage of $\{0, \dots, k-1\}^*$ of the form:

$$u_1 v_1^* u_2 v_2^* \dots u_p v_p^* \Sigma_{\ell,m,c},$$

where $p, \ell, m, c \in \mathbb{N}$, $u_1, \dots, u_p, v_1, \dots, v_p$ are finite words over the alphabet $[k] := \{0, \dots, k-1\}$, and $\Sigma_{\ell,m,c}$ is the set of all words of length a multiple of $\ell$ that are the base-$k$ expansion (*without* leading zeros) of a word congruent to $-c$ modulo $m$, where we take $\Sigma_{\ell,m,c}$ to be the empty language when $m = 0$ or $\ell = 0$. We suppose without loss of generality that $\Sigma_{\ell,m,c}$ does not include leading zeros, increasing $p$ if necessary so that any string of zeros preceding an element of $\Sigma_{\ell,m,c}$ is absorbed into $v_p^*$.

Theorem 2.15 allows us to observe the following fact.

***Fact 2.19*** [18] *A one-dimensional set* $X \subseteq \mathbb{N}$ *is definable in* $(\mathbb{N}, +, k^{\mathbb{N}})$ *if and only if* $X = [L]_k$, *where* $L$ *is a finite union of* $k$*-Semenov-type languages. In particular, if* $U$ *and* $V$ *are, respectively, a sparse regular sublanguage of* $\{0, \dots, k-1\}^*$ *and a sublanguage of* $\{0, \dots, k-1\}^*$ *consisting of base-*$k$ *expansions (without leading zeros) of a set definable in* $(\mathbb{N}, +, k^{\mathbb{N}})$, *then* $[UV]_k$ *is definable in* $(\mathbb{N}, +, k^{\mathbb{N}})$.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

## 3 Proof of the dichotomy for cycle languages

In this section, we'll prove the following result.

***Theorem 3.1*** *Let $\mathcal{A} = (Q, \{q_0\}, \Sigma_k, \delta, W)$ be a DFA which reads strings left-to-right. If $p \in Q$ and $X = \{[w]_k : w \in L_{p\to p}\}$, then either $X$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$ or $V_k$ is definable in $(\mathbb{N}, +, X)$.*

We give an overview of how we shall prove Theorem 3.1. We show that if $X$ is a set of natural numbers that is not Presburger definable and that corresponds to a cycle language for an automaton with input alphabet $\Sigma_k$ for some $k$, then (after replacing $k$ by a power if necessary) we may assume that there is a distinguished letter $a \in \{0, \ldots, k-1\}$ such that we can define a function $F : X \to k^{\mathbb{N}}$, which is a very close definable approximation to the function $V_{k,a}|_X$, where $V_{k,a}$ is the map that takes a natural number $n$ as input and outputs $k^i$, where $i$ is the unique nonnegative integer with the property that the base-$k$ expansion of $n$ ends with exactly $i$ copies of the digit $a$.

Having defined this map $F$, we show that we can extend $F$ to a definable approximation of the function $V_{k,a}$ on $X_{\leq P}$, where $P$ is a natural number and $X_{\leq P}$ is the set of all numbers that are sums of at most $P$ elements of $X$. Now we use a theorem of Bell et al. [5] to show that if $X$ is not sparse, then we can define an approximation of $V_{k,a}$ on all of $\mathbb{N}$. From here, we can show that one can define $V_k$ and so we deduce that one can define all $k$-automatic sets from $X$.

To prove Theorem 3.1, we may assume that $X$ is the set of base-$k$ expansions of a non-empty cycle language, and hence there is some word $w$ such that $\delta(p, w) = \{p\}$ for a fixed state $p \in Q$. We henceforth let

$$L := L_{p\to p} = \{w : \delta(p, w) = \{p\}\}, \tag{1}$$

and so $X = [L]_k$.

Since $L$ is non-empty, there is some non-empty word $w \in \Sigma_k^*$ such that $\delta(p, w) = \{p\}$. Possibly by replacing $k$ by a power, we may assume without loss of generality that $w$ is a letter $a \in \{0, 1, \ldots, k-1\}$. Furthermore, the pigeonhole principle gives that there is some positive integer $i$ such that $\delta(q, 0^i) = \delta(q, 0^{2i})$ and $\delta(q, (k-1)^i) = \delta(q, (k-1)^{2i})$ for all $q \in Q$. Thus, we may replace $k$ by a power and henceforth assume the following hold:

(i) There is a distinguished letter $a \in \{0, \ldots, k-1\}$ with the property that

$$\delta(p, a) = \{p\}. \tag{2}$$

(ii) We have $\delta(q, 00) = \delta(q, 0)$ for all $q \in Q$.
(iii) We have $\delta(q, (k-1)(k-1)) = \delta(q, k-1)$ for all $q \in Q$.
(iv) We also assume that $\mathcal{A}$ is minimal.

We shall say that a word $w \in \Sigma^*$ is *idempotent* (with respect to the automaton $\mathcal{A}$) if for all states $q \in Q$, we have $\delta(q, ww) = \delta(q, w)$. That is, applying the transition corresponding to $w$ twice has the same effect as applying it once. In particular, in our setting, we assume that the words 0 and $k-1$ are idempotent in this sense.

We now let $f$ denote the characteristic function of the set $X$. Then $f : \mathbb{N} \to \{0, 1\}$ is $k$-automatic and hence there are finitely many maps $f = f_1, \ldots, f_s$ such that, for each

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

$c \geq 0$ and each $j \in \{0, \ldots, k^c - 1\}$, we have $f(k^c n + j) = f_i(n)$ for some $i \in \{1, \ldots, s\}$, which depends on $c$ and $j$, by [2, Theorem 6.6.2]. We call these maps $f_1, \ldots, f_s$, the *kernel* of the map $f$. We may assume without loss of generality that $X$ is neither sparse nor eventually periodic. In particular, since $X$ is not Presburger definable, $(\mathbb{N}, +, X)$ defines the set $k^{\mathbb{N}}$ by Fact 2.17. In addition, there exists some natural number $b$ such that each $f_i$ can be realized as $f(k^r n + j)$ for some $r, j$ with $r \leq b$.

For the remainder of the proof, we divide the proof into three cases:

(I): $\delta(p, 0) = \{p\}$;
(II): $\delta(p, 0) = \varnothing$ or $\delta(p, 0) = \{p'\}$, where $p'$ is in a different strongly connected component from $p$;
(III): $\delta(p, 0) = \{p'\}$, where $p'$ and $p$ are in the same strongly connected component.

We note that, in Case (I), we can take $a = 0$ and this simplifies the proof of this case somewhat, but we will handle cases (I) and (II) simultaneously.

## 3.1 Proof in cases (I) and (II)

Throughout this section, we assume the following:

- One of cases (I) or (II) above holds.
- $X$ is not definable in $(\mathbb{N}, +, k^{\mathbb{N}})$.

It follows that $k^{\mathbb{N}}$ is definable in $(\mathbb{N}, +, X)$ by Fact 2.17.

We now define a series of functions $F_R : X \to k^{\mathbb{N}}$ for $R \in \mathbb{N}$ as follows. If $0 \in X$, we take $F_R(0) = 1$, and for $n \geq 1$ in $X$, we take $F_R(n)$ to be the largest $k^i \in k^{\mathbb{N}}$ such that $k^{i-1} \leq n$ and for all $0 \leq r \leq R$ and for all words $v$ with $|v| = j \leq r + i$, we have

$$k^r n - [a^{j-r}0^r]_k + [v]_k \in X \iff v \in L \qquad \text{if } j > r, \tag{3}$$

$$k^r n + [a^{r-j}v]_k \in X \iff v \in L \qquad \text{if } j \leq r. \tag{4}$$

We note two things about $F_R$:

- For $n \geq 1$, $k^0$ (i.e., $i = 0$) has the property that for all $r \geq 0$ and all words of length at most $r$ with $|v| \leq r$, the condition in Equation (4) holds, and so $F_R(n) \geq 1$ for all $n$. Therefore, $\min_R F_R(n)$ is well-defined; we denote this $F(n)$.
- Note that because $F_R$ is defined as the largest $k^i$ such that a universal quantification over $r \leq R$ holds, we must have that $(F_R(n))_R$ is monotone decreasing (in $R$).

We will now show that, due to the pumping lemma, the point at which the sequence $(F_R(n))_R$ stabilizes is uniform in $n$.

***Lemma 3.2*** *Let $\mathcal{A}'$ be a deterministic finite automaton recognizing $L$ with set of states $Q$, and let $M = (|Q| + 1)^2$. Then $F_M$ and $F$ are identical.*

**Proof** Fix $n$. By the above remarks, this is equivalent to the claim that for all $R > M$, $F_M(n) = F_R(n)$.

We proceed by induction on $R$. Assume that $F_M(n) = F_{M+1}(n) = \cdots = F_{R-1}(n)$. We need to show that $F_{R-1}(n) = F_R(n)$. Let $F_{R-1}(n) = k^i$; it suffices to show that for $r = R$ and for all words $v$ with $|v| = j \leq r + i$, the appropriate condition (either (3) or (4)) is satisfied.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

We first consider the case where $|v| \leq r$. In this case, we need to show that $k^r n + [a^{r-|v|}v]_k \in X \iff v \in L$. Let $[u]_k = n$; then equivalently, we want $ua^{r-|v|}v \in L \iff v \in L$.

Let $L' = \{a^{r'-|v'|}v' : ua^{r'-|v'|}v' \in L\}$. By Lemma 2.9, the automaton for $L'$ has at most $M$ states.

Consider the string $a^{r-|v|}v$. Note that this string has length $r = R > M$, so by the pumping lemma, we may write $a^{r-|v|}v = xyz$, where $y$ is nonempty and each $xy^m z \in L' \iff xyz \in L'$. Then $xz = a^{r'-|v'|}v'$ for some $r' < r$. We know that $ua^{r'-|v'|}v' \in L \iff v \in L$ by inductive assumption, so $ua^{r-|v|}v \in L \iff xyz \in L' \iff xz \in L' \iff ua^{r'-|v'|}v' \in L \iff v \in L$.

Now we consider the case where $|v| > r$. In this case, we need to show that $k^r n - [a^{j-r}0^r]_k + [v]_k \in X \iff v \in L$. Let $v = v_0 v_1$, where $|v_1| = r$; then the condition $k^r n - [a^{j-r}0^r]_k + [v]_k \in X$ may be rewritten as $k^r(n - [a^{j-r}]_k + [v_0]_k) + [v_1]_k$. Let $w_{j,v_0}$ be a word such that $[w_{j,v_0}]_k = n - [a^{j-r}]_k + [v_0]_k$; then we want to show that $w_{j,v_0}v_1 \in L \iff v_0 v_1 \in L$.

In this case, we set $L'_{j,v_0} = \{v_1 : w_{j,v_0}v_1 \in L \leftrightarrow v_0 v_1 \in L\}$. By Lemmas 2.9 and 2.10, the automaton for $L'_{j,v_0}$ has at most $M$ states. Because $|v_1| = r = R > m$, we apply the pumping lemma and let $v_1 = xyz$ with $y$ nonempty and $xy^m z \in L'_{j,v_0} \iff xyz \in L'_{j,v_0}$ for all $m$, in particular, $xz \in L'_{j,v_0} \iff xyz \in L'_{j,v_0}$.

Let $v'_1 = xz$, and let $|v'_1| = r' < r$. Let $j' = j + r' - r$. Then:

$$\begin{aligned}
[w_{j,v_0}v'_1]_k &= k^{r'}(n - [a^{j-r}]_k + [v_0]_k) + [v'_1]_k \\
&= k^{r'}n - k^{r'}[a^{j-r}]_k + k^{r'}[v_0]_k + [v'_1]_k \\
&= k^{r'}n - k^{r'}[a^{j-r}]_k + [v_0 v'_1]_k \\
&= k^{r'}n - k^{r'}[a^{j'-r'}]_k + [v_0 v'_1]_k \\
&= k^{r'}n - [a^{j'-r'}0^{r'}]_k + [v_0 v'_1]_k.
\end{aligned}$$

By inductive assumption, $k^{r'}n - [a^{j'-r'}0^{r'}]_k + [v_0 v'_1]_k \in X \iff v_0 v'_1 \in L$, that is, $w_{j,v_0}v'_1 \in L \iff v_0 v'_1 \in L$. Therefore, $xz = v'_1 \in L'_{j,v_0}$. It follows that $xyz = v_1 \in L'_{j,v_0}$, so $w_{j,v_0}v_1 \in L \iff v_0 v_1 \in L$, as required. ∎

Intuitively, $F$ is trying to capture (imperfectly) the length of the largest suffix of the base-$k$ expansion that consists entirely of $a$'s. We note that using geometric series, it is not difficult to show that the limiting map $F$ is definable in $(\mathbb{N}, +, X)$.

***Lemma 3.3*** *Under the assumptions above, the map $F$ is definable in $(\mathbb{N}, +, X)$.*

**Proof** We will construct, step-by-step, several definable functions and relations that ultimately give us $F_R$ for a fixed $R$ when put together. As $F = F_M$ by the previous lemma, we conclude that $F$ itself is definable. Throughout this proof, we work in the expanded language $\{+, -, \equiv_n, (x \mapsto nx)_{n\in\mathbb{N}}, <, X, k^{\mathbb{N}}\}$. We do this without loss of generality because it is elementary to show that the relations $<, >, \leq, \geq, \neq$, modular equivalence $\equiv_n$ modulo a constant $n$, multiplication $nx$ by a constant $n$, the graph of the partial function $-$, and all natural number constants themselves are definable

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

in Presburger arithmetic, and because Fact 2.17 guarantees that we can define $k^{\mathbb{N}}$ in $(\mathbb{N}, +, X)$.

- Let $\ell(x)$ be the smallest $k^i \in k^{\mathbb{N}}$ such that $x < k^i$. The graph of this function is defined by the formula $y \in k^{\mathbb{N}} \wedge \forall p_i \in k^{\mathbb{N}}\ (p_i > x \leftrightarrow p_i \geq y)$. Note that if $v$ starts with no leading zero, $\ell([v]_k)$ is encoded by the word $10^{|v|}$.
- Let $L(x, y)$ be the binary relation that holds precisely when there is a word $v$ with length $j$ such that $[v]_k = x$, $k^j = y$, and $v \in L$. The definition of this relation depends on whether we are in case (I) or (II). If case (I) holds, then for all $v$, we have $v \in L \iff 0v \in L$. Therefore, in this case, $L(x, y)$ is defined by the formula $y \geq \ell(x) \wedge x \in X$. If case (II) holds, then no word beginning with 0 is in $L$. Therefore, in this case, $L(x, y)$ is defined by the formula $y = \ell(x) \wedge x \in X$.
- Let $a(x, y)$ be the partial function where, if $x = k^i$ and $y = k^j$, then $a(x, y) = [a^{i-j}0^j]_k$. For example, $a(7, 4) = [aaa0000]_k$. Notice that $a(i, j) = [a]_k(k^j + \cdots + k^{i-1}) = [a]_k \frac{k^i - k^j}{k-1}$. Because $[a]_k$ and $(k-1)$ are constants, we therefore define the graph of $a$ by the formula $x \in k^{\mathbb{N}} \wedge y \in k^{\mathbb{N}} \wedge (k-1)z = [a]_k(x - y)$.
- Fix an arbitrary constant $r$. Let the relation $E_{(3),r}(x, y, n)$ hold if there is a word $v$ with length $j$ such that $[v]_k = x$, $k^j = y$, and equivalence (3) holds. Note that $k^r$ is a constant here. Then $E_{(3),r}$ is defined by the formula $y \geq \ell(x) \wedge (k^r n + x - a(y, k^r) \in X \leftrightarrow L(x, y))$.
- Fix an arbitrary constant $r$. Let the relation $E_{(4),r}(x, y, n)$ hold if there is a word $v$ with length $j$ such that $[v]_k = x$, $k^j = y$, and equivalence (4) holds. Note that $k^r$ is a constant here. Then $E_{(4),r}$ is defined by the formula $y \geq \ell(x) \wedge (k^r n + a(k^r, y) + x \in X \leftrightarrow L(x, y))$.
- Fix an arbitrary constant $r$. Let the relation $E_r(x, y, n)$ hold if there is a word $v$ with length $j$ such that $[v]_k = x$, $k^j = y$, and equivalence (3) or (4) holds depending on whether $j > r$. Then $E_r$ is defined by the formula $(y > k^r \wedge E_{(3),r}(x, y, n)) \vee (y \leq k^r \wedge E_{(4),r}(x, y, n))$.
- Let the relation $A_r(n, z)$ hold if $z = k^i$, $k^{i-1} \leq n$, and for all $v$ with $|v| = j \leq r + i$, we have $E_r(x, y, n)$ when $[v]_k = x$ and $k^j = y$. Then $A_r$ is defined by the formula $z \in k^{\mathbb{N}} \wedge \ell(n) \geq z \wedge \forall x\, \forall y\, ((y \in k^{\mathbb{N}} \wedge \ell(x) \leq y \wedge y \leq k^r z) \rightarrow E_r(x, y, n))$.
- Let the relation $A_{\leq R}(n, z)$ hold if $A_r(n, z)$ holds for all $r \leq R$. Then for a fixed $R$, $A_{\leq R}$ is defined by the formula $A_0(n, z) \wedge A_1(n, z) \wedge \cdots \wedge A_R(n, z)$.
- The function $F_R(n)$ returns the largest $z = k^i \in k^{\mathbb{N}}$ such that $k^{i-1} < n$ (equivalently $k^i < kn$) and such that $A_{\leq R}(n, k^i)$ holds. Then the graph of $F_R$ is defined by the formula $z \in k^{\mathbb{N}} \wedge z < kn \wedge A_{\leq R}(n, z) \wedge \forall z'\, ((z' \in k^{\mathbb{N}} \wedge z' < kn \wedge A_{\leq R}(n, z')) \rightarrow z' \leq z)$. ∎

***Example 1*** We give an example to show how to compute the function $F$ in a special case. Consider the case when $k = 3$ and when our distinguished letter is $a = 1$ and when $X$ is the set of natural numbers whose base-$k$ expansion is in the cycle language of the state $q_2$ of the automaton in Figure 1, where we again read strings from left-to-right. We claim that $F(22) = F([211]_3) = 3$.

Then if we take $n = [211]_3 = 22$, $F(n) = 3$. To see this, notice that if $r \geq 0$ and $v$ is a word with length at most $r + 1$ with $|v| = r + 1$, then $3^r n - [1^{|v|-r}0^r]_3 + [v]_3$ has ternary

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

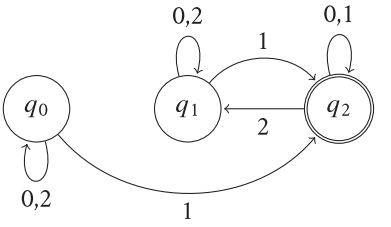



*Figure 1*: An automaton with cycle language at $q_2$ having $F(22) = 3$.

expansion $21v$, which is in the cycle language of the state $q_2$ if and only if $v$ is. If $|v| \leq r$, then

$$3^r n + [1^{r-j}v]_k$$

has ternary expansion $21^{2+r-|v|}v$, which is again in the cycle language if and only if $v$ is. It follows that $F(22) \geq 3$. To see that $F(22) < 9$, observe that for $r = 0$, $v = 00$, then $n - [11]_3 + [v]_3 = [200]_3$, which is not in $X$, while $[00]_3$ is. Thus, $F(22) = 3$, as claimed.

We now return to the proof of Theorem 3.1 in cases (I) and (II). By definition of the map $F$, we also have implication:

$$n \leq k^i \implies F(n) \leq k^{i+1}, \tag{5}$$

for $n \in X$.

As mentioned earlier, we view the function $F$ as a (definable) proxy for the function $V_{k,a}$ on the set $X$, where $V_{k,a}(n) = k^i$ if the base-$k$ expansion of $n$ has a suffix of length $i$ consisting entirely of $a$'s and no longer suffix with this property. As it turns out, we can show that $F$ approximates $V_{k,a}$ up to multiplicative constants.

***Lemma 3.4*** *There is a natural number $\beta \geq 1$, depending only on $X$ such that, for $n \in X$, we have $F(n) \geq V_{k,a}(n)/k^\beta$.*

**Proof** Since there are only finitely many distinct vectors of sets $(\delta(q, a^r))_{q\in Q}$ for $r \in \mathbb{N}$, by the pigeonhole principle, there exist natural numbers $r_1, r_2$, with $r_1 < r_2$, such that $\delta(q, a^{r_1}) = \delta(q, a^{r_2})$ for all $q \in Q$. We claim that we can take $\beta = r_2$.

If $n = 0$, there is nothing to prove, so we may assume that $n \in X$ is positive. In addition, if the base-$k$ expansion of $n$ is of the form $a^m$ for some $m \geq 1$, then it follows from the definition that $F(n) \geq V_{k,a}(n)/k = k^{m-1}$. Thus, we may assume that $n$ has at least one base-$k$ digit not equal to $a$.

Let $wja^m$ denote a base-$k$ of $n$ that is in $L$, where $w \in \Sigma^*$ is a possibly empty word and $j \in \Sigma_k \backslash \{a\}$. We claim that $F(n) \geq k^{m-r_2}$. If $m \leq r_2$, the claim is clear, since $F(n) \geq 1$ for all $n \in X$, so we assume that $m > r_2$.

Since $\delta(p, wja^m) = \{p\}$, by definition of $r_1$ and $r_2$, there exists $m_0 < r_2$ such that $\delta(p, wja^{m_0}) = \{p\}$. Moreover, since $\delta(p, a) = \{p\}$, we in fact have $\delta(p, wja^e) = \{p\}$ for all $e \geq m_0$.

To show that $F(n) \geq V_{k,a}(n)/k^{r_2} = k^{m-r_2}$, let $r \geq 0$ and let $v \in \Sigma_k^*$ be a word of length at most $m - r_2 + r$

We must show that the conditions in Equations (3) and (4) hold.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

First, if $|v| > r$, then $k^r n - [a^{|v|-r}0^r]_k + [v]_k$ has base-$k$ expansion of the form $wja^e v$ for some $e \geq m_0$ and since $wja^e$ is in $L$, we see that $wja^e v \in L$ if and only if $v \in L$. Thus, since we are in either case (I) or (II), we see that

$$k^r n - [a^{|v|-r}0^r]_k + [v]_k \in X \iff v \in L$$

in this case.

Next, if $|v| \leq r$, then $k^r n + [a^{r-|v|}v]_k$ has base-$k$ expansion $wja^e v$ with $e \geq m_0$ and so as in the previous case, we have that $k^r n + [a^{r-|v|}v]_k \in X \iff v \in L$.

Thus, $F(n) \geq k^{m-r_2} = V_{k,a}(n)/k^\beta$. ∎

We now have shown that $F(n)/V_{k,a}(n)$ is bounded below by $1/k^\beta$ for some $\beta$ that depends only on $X$ and not on $n$. We will next show that this ratio is bounded above, although this requires our assumption that the set $X$ not be eventually periodic.

We first require a monotonicity result.

***Lemma 3.5*** *For $w \in L$ and all $i \geq 1$, $F([wa^i]_k) \geq k^i F([w]_k)$.*

**Proof** Since $w \in L \implies wa \in L$, it suffices to prove the result when $i = 1$ and the general result will then follow by induction.

Let $t$ be such that $k^t = F(n)$, where $n = [w]_k$ with $w \in L$. Then Equations (3) and (4) give that for all $r \geq 0$ and all words $v$ of length at most $r + t$ with $|v| = j \leq r + t$, we have

$$k^r n - [a^{j-r} \cdot 0^r]_k + [v]_k \in X \iff v \in L$$

if $j > r$; and we have

$$k^r n + [a^{r-j}v]_k \in X \iff v \in L$$

if $j \leq r$.

Now let $m = kn + a$, so $m = [wa]_k$. Then for $r \geq 0$ and a word $v$ of length at most $r + t + 1$, if $|v| > r$, then

$$\begin{aligned} k^r m - [a^{|v|-r}0^r] + [v]_k &= k^{r+1}n + ak^r - [a^{|v|-r}0^r]_k + [v]_k \\ &= k^{r+1}n - [a^{|v|-(r+1)}0^{r+1}]_k + [v]_k, \end{aligned}$$

and so we see that this is in $X$ if and only if $v \in L$.

Similarly, if $|v| \leq r$, then

$$k^r m + [a^{r-j}v]_k = k^{r+1}n + [a^{r+1-j}v]_k,$$

and so this is $X$ if and only if $v \in L$. It follows that $F(kn + a) = F([wa]_k) \geq kF([w]_k)$, and the general result now follows. ∎

We now give an argument that shows that when $F(n)$ takes some value, we can find an $n'$ in $X$ such that $F(n') \geq F(n)$ with the property that $n'$ is bounded above by $F(n')$ up to a multiplicative constant.

***Lemma 3.6*** *Let $n \in X$. Then there is a fixed natural number $\alpha$, depending only on $X$, such that whenever $F(n) = k^y > V_{k,a}(n)$, there exists $n' \in X$ with $F(n') \geq F(n)$, $V_{k,a}(n) = V_{k,a}(n')$, and $n' < k^\alpha F(n')$.*

**Proof** We claim we can take $\alpha := M + 2$, where $M$ is the number of states in our minimal automaton $\mathcal{A}$. By definition of $F(n)$, we have $n \geq k^{t-1}$ and so every base-$k$ expansion of $n$ has at least $t$ digits. We pick a base-$k$ expansion $wu$ of $n$ in $L$, where

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

$w, u \in \Sigma_k^*$ and $|u| = t$. When analyzing Equations (3) and (4), the main difficulty that arises is that carries can occur, either when performing additions or subtractions.

We first note that we may assume that the prefix $w$ is non-empty: if not, $n = [u]_k < k^{t+1}$ and so if we take $n' = n$, we have $n' < k^\alpha F(n) = k^\alpha F(n')$, so the result is clear in this case.

Thus, we assume that $w$ is non-empty and, in this case, it can be written uniquely in the form $w = w_0 c^\ell$ for some $\ell \geq 0$, where $c \in \{0, k-1\}$ and $w_0$ is either empty or a word whose last letter is in $\{1, 2, \ldots, k-2\}$.

Thus, we now write $n = [w_0 c^\ell u]_k$. We consider the cases when $w_0$ is empty and when it is not empty separately. To ease arguments, we write $w_1 \sim_L w_2$ to mean that either $w_1$ and $w_2$ are both in $L$ or neither is in $L$.

**Case A:** $w_0$ is not the empty word.

In this case, we write $w_0 = w_0' s$, where $s \in \{1, 2, \ldots, k-2\}$. Since our minimal automaton $\mathcal{A}$ has at most $M$ states, there is some word $w_0''$ of length $< M$ such that $\delta(q_0, w_0') = \delta(q_0, w_0'')$. We now take $n' = [w_0'' s c^{\min(\ell,2)} u]_k$. Notice that $n' < k^{t+M+2}$, so it suffices to show that $F(n') \geq k^t$.

It now suffices to verify that Equations (3) and (4) hold for words $v$ of length $t + r$ for $n'$. We only consider Equation (3), since Equation (4) is verified in a similar manner.

Since $F(n) = k^t$, by Equation (3), we have for $r \geq 0$ and $|v| \leq t + r$, if $|v| > r$, then

$$k^r n - [a^{j-r}0^r]_k + [v]_k = k^r [c^\ell u]_k - [a^{j-r}0^r]_k + [v]_k$$

is in $X$ if and only if $v \in L$. Moreover, observe that the quantity $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k$ is greater than or equal to $-k^{t+r} + 1$ and is at most $2 \cdot k^{r+t} - 1$.

We now look at various subcases.

**Subcase (i):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k < 0$, $c = k-1$, and $\ell \geq 2$. In this case,

$$k^r [w_0' s c^\ell u]_k - [a^{j-r}0^r]_k + [v]_k$$

either has base-$k$ expansion of the form $w_0' s c^{\ell-1}(c-1)v'$ or of the form $w_0' s c^\ell v'$ for some $v'$ which depends on $u$ and $v$ but not on $\ell \geq 2$. In particular, since $k-1$ is idempotent with respect to $\mathcal{A}'$ we see that $w_0' s c^{\ell-1}(c-1)v' \sim_L w_0'' s c(c-1)v'$ and $w_0' s c^{\ell-1} v' \sim_L w_0'' s c c v'$. Thus, we see that $k^r n' - [a^{j-r}0^r]_k + [v]_k \in X \iff v \in L$ in this case.

**Subcase (ii):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k < 0$, $c = k-1$, and $\ell < 2$.

This case is handled similarly to Subcase (i), but where we now use $n' = [w_0'' s c^\ell u]_k$.

**Subcase (iii):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k < 0$, $c = 0$, and $\ell \geq 2$. In this case,

$$k^r [w_0' s c^\ell u]_k - [a^{j-r}0^r]_k + [v]_k$$

either has base-$k$ expansion of the form $w_0'(s-1)(k-1)^\ell v'$ or of the form $w_0' s c^\ell v'$ for some $v'$ which depends on $u$ and $v$ but not on $\ell \geq 2$. Then, using the fact that $c = 0$ is idempotent with respect to $\mathcal{A}'$ we see as in Subcase (i) that if $n' = [w_0'' s 0^\ell]_k$, then $k^r n' - [a^{j-r}0^r]_k + [v]_k \in X \iff v \in L$.

**Subcase (iv):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k < 0$, $c = 0$, and $\ell < 2$.

This case is handled similarly to Subcase (ii), but where we now use $n' = [w_0'' s c^\ell u]_k$.

**Subcase (v):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k \geq 0$, $c = k-1$, and $\ell \geq 2$.

In this case,

$$k^r [w_0' s c^\ell u]_k - [a^{j-r}0^r]_k + [v]_k$$

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

either has base-$k$ expansion of the form $w_0'(s+1)0^\ell v'$ or of the form $w_0's(k-1)^\ell v'$ for some $v'$ which depends on $u$ and $v$ but not on $\ell \geq 2$. In this case, we argue as in Subcase (i) with $n' = [w_0'' sc^2 u]_k$.

**Subcase (vi):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k \geq 0$, $c = k-1$, and $\ell < 2$.

Here, we argue similarly in Subcase (iv) with $n' = [w_0'' sc^\ell u]_k$.

**Subcase (vii):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k \geq 0$, $c = 0$, and $\ell \geq 2$.

In this case,

$$k^r[w_0' sc^\ell u]_k - [a^{j-r}0^r]_k + [v]_k$$

either has base-$k$ expansion of the form $w_0's0^{\ell-1}1v'$ or of the form $w_0's0^\ell v'$ for some $v'$ which depends on $u$ and $v$ but not on $\ell \geq 2$. Then we argue as in the earlier subcases with $n' = [w_0'' s0^2 u]_k$.

**Subcase (viii):** $k^r[u]_k - [a^{j-r}0^r]_k + [v]_k \geq 0$, $c = 0$, and $\ell < 2$.

This is done as in the previous subcase with $n' = [w_0'' s0^\ell u]_k$.

**Case B:** $w_0$ is the empty word. This case is done similarly to Case A, where we take $n' = [c^{\min(\ell,2)}u]_k$. Thus, we have found $n' \in X$ with $F(n') \geq k^t$ and $n' \leq k^{M+2}k^t$. ∎

***Definition 3.7*** We say that two positive integers $n_1, n_2$ are $(k,a)$-*equivalent* if there is some $i \geq 0$ such that either $n_1 = k^i n_2 + (k^i - 1)a/(k-1)$ or $n_2 = k^i n_1 + (k^i - 1)a/(k-1)$.

This notion of $(k,a)$-equivalence is easily seen to be an equivalence relation, and intuitively, it says that $n_1$ and $n_2$ are equivalent if the base-$k$ expansion of one can be obtained from the other by either appending a suffix of $a$'s at the end or by deleting a string of $a$'s at the end.

In the following lemma, we will use Mahler series to prove that an automatic set that is periodic on a sufficiently long interval is necessarily eventually periodic. A power series $h(t) \in \mathbb{Z}[[t]]$ is *k-Mahler*, if it satisfies a nontrivial equation

$$\sum_{i=0}^{r} P_i(t)h(t^{k^i}) = 0 \tag{6}$$

with $P_0P_r \neq 0$ and $P_0, \ldots, P_r$ polynomials in $t$. If $h(t)$ is the characteristic function of a $k$-automatic set, then $h(t)$ is $k$-regular [1] and thus $k$-Mahler [4]. For a power series $f(t) \in \mathbb{Z}[[t]]$, we write $f(t) = O(t^N)$ to mean that $f(t) \in t^N\mathbb{Z}[[t]]$.

***Lemma 3.8*** *Let $T$ be a $k$-automatic set and let $C \geq 0$. Then there exists a natural number $N$ depending only on $T$ and $C$, such that if there is some $p \geq 1$ such that $n + p \in T \iff n \in T$ for $n \in [Cp, Np)$, then $T$ is eventually periodic.*

**Proof** Let $h(t) \in \mathbb{Z}[[t]]$ be the characteristic function of $T$. Then $h$ is $k$-Mahler and so it satisfies a nontrivial $k$-Mahler equation as given in Equation (6).

The condition that $j + p \in T \iff j \in T$ for $j + p \in [Cp, Np)$ says that $h(t)(1 - t^p) = q(t) + O(t^{(N-1)\ell})$ for some polynomial $q$ of degree $< (C+1)p$. Then

$$\sum_{i=0}^{r} P_i(t)q(t^{k^i})/(1 - t^{pk^i}) = O(t^{(N-1)\ell}).$$

Multiplying through by $1 - t^{pk^r}$, we then see

$$\sum_{i=0}^{r} P_i(t)q(t^{k^i})(1 - t^{pk^r})/(1 - t^{pk^i}) = O(t^{(N-1)p}).$$

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

Now let $M$ denote the maximum of the degrees of $P_0, \ldots, P_r$. Then the left-hand side has degree at most $M + k^r p(C+1)$. Thus, if $M + k^r p(C+1) < (N-1)p$, then the left-hand side must be identically zero. In particular, taking $N \geq M + k^r(C+2) + 1$ gives that the left-hand side vanishes. But now we claim that $h(t)$ must be equal to $q(t)/(1-t^\ell)$. To see this, we note that, if this is not the case, then we may write $h(t) = q(t)/(1-t^\ell) + t^e h_0(t)$ with $h_0(0) \neq 0$ and

$$e \geq (N-1)\ell \geq (M + k^L + 1)\ell.$$

Then since

$$\sum_{i=0}^{r} P_i(t) h(t^{k^i}) = 0$$

and

$$\sum_{i=0}^{r} P_i(t) q(t^{k^i})/(1 - t^{\ell k^i}) = 0,$$

we must have

$$\sum_{i=0}^{r} P_i(t) t^{ek^i} h_0(t^{k^i}) = 0.$$

But notice that all terms in the left-hand side other than $P_0(t)t^e h_0(t)$ vanish to order at least $ek$ at $t = 0$, while $P_0(t)t^e h_0(t)$ vanishes to order at most $t^e + \deg(P_0) \leq e + M$ at $t = 0$. Since $e + M < ke$, we get that $h(t) = q(t)/(1-t^p)$, and so $h(t)$ is a rational function whose coefficients lie in $\{0,1\}$ and so the sequence of coefficients of $h(t)$ is eventually periodic. Thus, $T$ is eventually periodic. ∎

We now show that when $X$ is not eventually periodic that $F(n)/V_{k,a}(n)$ is bounded above. This is the most technical part of the proof of Theorem 3.1 in cases (I) and (II) and is done via two technical lemmas.

***Lemma 3.9*** *Suppose that there are natural numbers $M$ and $P$ such that for all words $v, v'$ with $|v'| = |v| = p \leq M$ and $[v']_k = P + [v]_k$, we have*

$$v \in L \iff v' \in L.$$

*Then $x \in X \iff x + P \in X$ for all $x \in [kP, k^M - P)$.*

**Proof** Suppose that $v$ and $v'$ are words with no leading zeros such that $[v']_k = [v]_k + P$ and $|v|, |v'| \leq \log_k(\min(F(n_i)k^c, F(n_j)))$. We now consider cases (I) and (II) separately.

**When we are in Case (II):** In this case, let $v, v'$ be words with no leading zeros such that $[v']_k = [v]_k + P$ and $[v']_k = m$. If $i = |v| = |v'|$, then by construction,

$$v \in L \iff n_j + P - [a^{|v|}]_k + [v]_k \in X \iff n_j - [a^p]_k + [v']_k \in X \iff v' \in L$$

and since we are in case (II), we see that $[v]_k \in X \iff [v']_k \in X$. On the other hand, if $v$ and $v'$ don't have the same length, then we similarly have $0^{|v'|-|v|}v \in L \iff v' \in L$ and since we are in case (II), $0^{|v'|-|v|}v \notin L$ and hence $v' \notin L$ and so $[v']_k \notin X$ in this case. Thus, we always have that if $m \in X$ and $m > P$, then $m - P \in X$. Now suppose that $m - P \in X$ but $m \notin X$. Now pick the smallest positive integer $s$ such that $k^s > P$.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

We claim that $X \cap [k^s, k^M)$ is empty. To see this, suppose that this is not the case and pick the smallest $m \in X \cap [k^s, k^M)$. Since $m \in X \implies m - P \in X$ for $m > P$, we see that $m \in [k^s, k^s + P)$. Hence, $m - P < k^s$ and so if $v$ and $v'$ are base-$k$ expansions of $m - P$ and $m$, respectively, and $v'$ has no leading zeros, then if $|v| = |v'|$, then $v$ necessarily has a leading zero. And so from the argument above since $v \notin L$, we then have $v' \notin L$ and so $m \notin X$, a contradiction. Thus, $X \cap [k^s, k^M)$ is empty and since $k^s \leq kP$, $x \in X \iff x + P \in X$ for all $x \in [kP, k^M - P)$. Thus, the result follows when we are in Case II. **When we are in Case (I):**

Let $m < k^M - P$ be a positive integer and let $v, v'$ be words with $[v]_k = m$ and $[v']_k = m + P$. Since we are in case (I), we may pad $v$ with leading zeros and assume that $|v| = |v'|$. Then we see that

$$v \in L \iff n_j + P - [a^{|v|}]_k + [v]_k \in X \iff n_j - [a^p]_k + [v']_k \in X \iff v' \in L$$

and since we are in case (I), we see that $[v]_k \in X \iff [v']_k \in X$. Hence, we have that $m \in X$ if and only $m + P \in X$ for all $m < k^M - P$. So the result follows in this case. ∎

***Lemma 3.10*** *If $X$ is not eventually periodic, then $F(n)/V_{k,a}(n)$ is bounded above.*

**Proof** Suppose, toward a contradiction, that $F(n)/V_{k,a}(n)$ is not bounded above. Then there exists a sequence $n_1, n_2, n_3, \ldots$ in $X$ with $F(n_i)/V_{k,a}(n_i) > k^i$ for all $i$. By Lemma 3.6, we can replace each $n_i$ by a corresponding $n_i'$ and assume additionally that $n_i < k^\alpha F(n_i)$ for some fixed $\alpha$ that does not depend on $i$ and depends only on $X$.

In particular, by Lemma 3.5, we have

$$\frac{F\left(k^r n_i + \frac{k^r-1}{k-1}a\right)}{V_{k,a}\left(k^r n_i + \frac{k^r-1}{k-1}a\right)} \geq k^i \tag{7}$$

for all $r \geq 0$. Moreover, since $n_i < k^\alpha F(n_i) \in k^{\mathbb{N}}$, we have

$$k^r n_i + \frac{k^r - 1}{k-1}a < k^{r+\alpha}F(n_i).$$

Thus, by Equation (5), we see that

$$F\left(k^r n_i + \frac{k^r - 1}{k-1}a\right) \leq k^{r+\alpha+1}F(n_i)$$

and so

$$k^r F(n_i) \leq F(k^r n_i + (k^r - 1)a/(k-1)) \leq k^{r+1+\alpha}F(n_i) \tag{8}$$

for all $r \geq 0$.

In particular, if $n_i < n_j$ and $n_i$ and $n_j$ are $(k, a)$-equivalent, then we must have

$$n_j = n_i k^r + (k^r - 1)a/(k-1)$$

for some $r \geq 0$ and so from the above, we have

$$\begin{aligned} F(n_j)/V_{k,a}(n_j) &= F(k^r n_i + (k^r - 1)a/(k-1))/(k^r \cdot V_{k,a}(n_i)) \\ &\leq k^{\alpha+1+r}F(n_i)/(k^r \cdot V_{k,a}(n_i)) \\ &\leq k^{\alpha+1}F(n_i)/V_{k,a}(n_i). \end{aligned}$$

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

Thus, since $n_j \to \infty$ and $F(n_j)/V_{k,a}(n_j) \to \infty$, we may refine our sequence $(n_i)$ if necessary that we may assume that they are pairwise $(k,a)$-inequivalent. Now, for each positive integer $n$, we let $t(n)$ denote the unique natural number $t$ such that $k^{t-1} \leq n < k^t$. Then the sequence $(n_i/k^{t(n_i)})$ lies in the compact set $[1/k, 1]$ and so by the Bolzano–Weierstrass theorem, it has a convergent subsequence. In particular, since $X$ is not eventually periodic, by Lemma 3.8, there is a natural number $N$ such that whenever $p > 0$ is such that $x \in X \iff x + p \in X$ for all $x \in [pk, M)$, we necessarily have $M < Np$.

Choose $\epsilon \in (0, 1/2k)$ so that

$$2k(N+1)\epsilon < k^{-\alpha}/2 \tag{9}$$

we see that we may find $(k,a)$-inequivalent $n_i$ and $n_j$ in this subsequence (by assumption) such that

$$|n_i/k^{t(n_i)} - n_j/k^{t(n_j)}| < \epsilon, \tag{10}$$

and we may assume in addition that $n_i < n_j$ and that $n_i > (a+1)(1+k\epsilon)/\epsilon$. We let

$$c = t(n_j) - t(n_i).$$

Then, after multiplying both sides of Equation (10) by $k^{t(n_j)}$, we have

$$|n_i k^c - n_j| < k^{t(n_j)}\epsilon \leq k n_j \epsilon.$$

In particular,

$$k^c < (1 + k\epsilon) n_j / n_i \tag{11}$$

and

$$|n_i k^c + (k^c - 1)a/(k-1) - n_j| < k n_j \epsilon + a k^c < k n_j \epsilon + a(1+k\epsilon) n_j / n_i. \tag{12}$$

Now we let $P := n_i k^c + (k^c - 1)a/(k-1) - n_j$, which is bounded above by $k n_j \epsilon + a k^c$. We consider only the case when $P > 0$, with the case when $P < 0$ being handled similarly. Then since $F(n_i) \geq k^i$, we see that $F(k^c n_i + (k^c - 1)a/(k-1)) \geq k^{i+c}$ by Lemma 3.5. Thus, using Equations (3) and (4) with $r = 0$, we see that for all words $v$ with $|v| = p \leq \log_k(\min(F(n_i)k^c, F(n_j)))$, we have

$$(k^c n_i + (k^c - 1)a/(k-1)) - [a^p]_k + [v]_k \in X \iff v \in L.$$

Thus, by our definition of $P$, we have

$$(n_j + P) - [a^p]_k + [v]_k \in X \iff v \in L$$

for $|v| = p \leq \log_k(\min(F(n_i)k^c, F(n_j)))$. But by definition of $n_j$, we also have

$$n_j - [a^p]_k + [v]_k \in X \iff v \in L$$

for $|v| = p \leq \log_k(\min(F(n_i)k^c, F(n_j)))$. It follows that if $v, v'$ are words of the same length $p \leq \log_k(\min(F(n_i)k^c, F(n_j)))$ with $[v']_k = [v]_k + P$ then

$$v \in L \iff n_j + P - [a^{|v|}]_k + [v]_k \in X \iff n_j - [a^p]_k + [v']_k \in X \iff v' \in L.$$

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

It now follows from Lemma 3.9 that $x \in X \iff x + P \in X$ for all $x \in [kP, \min(F(n_i)k^c, F(n_j)) - P)$. Observe that, by Equation (12), we have

$$P \cdot (N+1) = (k^c n_i + (k^c - 1)a/(k-1) - n_j) \cdot (N+1) < (kn_j\epsilon + a(1+k\epsilon)n_j/n_i) \cdot N.$$

Recall that we refined our sequence of $n_\ell$'s earlier and so we have $n_i > (a+1)(1+k\epsilon)/\epsilon$ and thus we have $P(N+1) < 2kn_j(N+1)\epsilon$.

Now, by our choice of $n_\ell$'s, we have $F(n_j) > k^{-\alpha} n_j$ and, by our choice of $\epsilon$, we then have $2k(N+1)\epsilon n_j < F(n_j)$. We similarly have

$$k^c F(n_i) > k^{c-\alpha} n_i > k^\alpha (n_j - kn_j\epsilon) < k^{-\alpha} n_j/2,$$

since $\epsilon < 1/2k$. Thus, we again have $2k(N+1)\epsilon n_j < k^c F(n_i)$. Hence, $P(N+1) < \min(F(n_j), k^c F(n_i))$. It follows that $x \in X \iff x + P \in X$ for $x \in [kP, PN)]$, but by our choice of $N$, this cannot hold since $X$ is not eventually periodic. ∎

We now show that $V_k$ can be defined using the map $F$, continuing with the notation introduced above and under the assumption that $X$ is neither Presburger definable nor sparse.

By Lemmas 3.4 and 3.10, the ratio $F(n)/V_{k,a}(n)$ is bounded above and below for $n \in X$, and hence there exists a finite set of integers $\{i_1, \ldots, i_r\}$ such that

$$F(n)/V_{k,a}(n) \in \{k^{i_1}, \ldots, k^{i_r}\} \tag{13}$$

for all $n \in X$.

Define $\tilde{X}$ as follows:

$$n \in \tilde{X} \iff n + \frac{k^t - 1}{k-1} a \in X, \tag{14}$$

where $t$ is the largest natural number such that $k^t \leq n$. We note that we can define $X$ from $\tilde{X}$ via the rule

$$m \in X \iff m - \frac{k^s - 1}{k-1} a \in \tilde{X}, \tag{15}$$

where $s$ is the unique nonnegative integer such that

$$k^s + \frac{k^s - 1}{k-1} a \leq m < k^{s+1} + \frac{k^s - 1}{k-1} a.$$

We note that the map from $\tilde{X}$ to $X$ and the map from $X$ to $\tilde{X}$ described in Equations (14) and (15) are inverses of one another and so these maps give bijections between the sets.

We define a map $\tilde{F} : \tilde{X} \to k^{\mathbb{N}}$ by

$$\tilde{F}(n) = F\left(n + \frac{k^t - 1}{k-1} a\right), \tag{16}$$

where $t$ is again the largest natural number such that $k^t \leq n$.

Both $\tilde{X}$ and $\tilde{F}$ are definable in the structure $(\mathbb{N}, +, X, k^{\mathbb{N}})$ by construction, and since $X$ is $k$-automatic but not Presburger definable, it follows from Fact 2.17 that $k^{\mathbb{N}}$ is definable in $(\mathbb{N}, +, X)$. Thus, to complete the proof of Theorem 3.1, we assume

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

that $X$ is neither sparse nor eventually periodic and show that $V_k$ is definable using $\tilde{X}$ and $\tilde{F}$.

***Lemma 3.11*** *Adopt the notation above and suppose that $X$ is neither sparse not eventually periodic. Then the following hold:*

(i) *$\tilde{F}(n)/V_k(n)$ is bounded above and below by constants for $n \in \tilde{X}$;*
(ii) *$\tilde{X}$ is not sparse;*
(iii) *if $n \in \tilde{X}$, then $kn \in \tilde{X}$;*
(iv) *$\tilde{F}(kn) \geq k \cdot \tilde{F}(n)$ for $n \in \tilde{X}$.*

**Proof** Let $n \in \tilde{X}$ and let $t$ be the unique natural number such that $k^t \leq n < k^{t+1}$. We take $m := n + \frac{k^t-1}{k-1}a$ and note that $m$ lies in $X$ by definition of $\tilde{X}$. By definition,

$$\tilde{F}(n) = F(m).$$

Notice that if the base-$k$ expansion of $n$ ends with precisely $s$ zeros, then the base-$k$ expansion of $m$ ends with exactly $s$ copies of $a$. Hence,

$$V_{k,a}(m) = V_k(n).$$

Since $F(m)/V_{k,a}(m)$ is bounded above and below for $m \in X$ by Lemmas 3.4 and 3.10, we deduce that $\tilde{F}(n)/V_k(n)$ is also bounded above and below. This proves part (i).

To prove part (ii), we note that, by Equation (15), we have a surjective map $h$ from the positive elements of $X$ to the positive elements of $\tilde{X}$ given by

$$m \mapsto m - \frac{k^s - 1}{k-1}a,$$

where $s$ is the unique integer satisfying

$$k^s + (k^s - 1)a/(k-1) \leq m < k^{s+1} + (k^s - 1)a/(k-1).$$

As noted by Equations (14) and (15), the map $h$ is injective and since $h(m) \leq m$, we see that the number of elements in $\tilde{X}$ up to a positive integer $x$ is at least as large as the number of elements in $X$ up to $x$, and so $\tilde{X}$ is not sparse since $X$ is not sparse.

For part (iii), suppose $n \in \tilde{X}$, so that $m = n + \frac{k^t-1}{k-1}a \in X$. Then

$$kn + \frac{k^{t+1}-1}{k-1}a = k\left(n + \frac{k^t-1}{k-1}a\right) + a = km + a.$$

Since $m \in X$ and $X$ is closed under the map $n \mapsto kn + a$, it follows that $km + a \in X$, and hence $kn \in \tilde{X}$. This proves (iii).

Finally, for part (iv), note from the above that

$$\tilde{F}(n) = F(m), \quad \tilde{F}(kn) = F(km + a).$$

Since $F(kn + a) \geq kF(n)$ holds for all $n \in X$ and $m \in X$, we conclude that

$$\tilde{F}(kn) = F(km + a) \geq kF(m) = k\tilde{F}(n).$$

This completes the proof of part (iv). ∎

***Lemma 3.12*** *Adopt the notation above. If $X$ is neither sparse nor eventually periodic, then there exists a subset $Y \subseteq \mathbb{N}$ and a map $H : Y \to k^{\mathbb{N}}$ with the following properties:*

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

(i) *$Y$ and $H$ are both definable in $(\mathbb{N}, +, X)$;*
(ii) *for $n \in Y$, $H(n)/V_k(n)$ takes only finitely many distinct values and $H(n) \leq V_k(n)$ for all $n \in Y$;*
(iii) *the set $Y_0$ of $n \in Y$ for which $H(n) = V_k(n)$ is not sparse and contains* 1*;*
(iv) *if $n \in Y_0$, then $kn \in Y_0$.*

**Proof** By Lemma 3.11, there exists a non-sparse set $\tilde{X}$ with a definable map $\tilde{F} : \tilde{X} \to k^{\mathbb{N}}$ such that $\tilde{F}(n)/V_k(n)$ is bounded above and satisfies properties (iii) and (iv) in the statement of the lemma. In particular, $\tilde{F}(n)/V_k(n)$ assumes finitely many values for $n \in X$, say $k^{p_1}, \ldots, k^{p_e}$ with $p_1 < p_2 < \cdots < p_e$. We let

$$T_j := \{n \in \tilde{X} : \tilde{F}(n)/V_k(n) = k^{p_j}\}.$$

Then since $\tilde{X}$ is not sparse, at least one set $T_i$ is not sparse and we let $j$ denote the largest index for which $T_j$ is not sparse. Then $T_i$ is sparse for $i > j$ and hence definable in $(\mathbb{N}, +, X)$, since $X$ is not Presburger definable. We now take

$$T_i' := \{n \in \mathbb{N} : \exists m, k^m n \in T_i\}.$$

Then $T_i'$ is again sparse for $i > j$ and therefore definable in $(\mathbb{N}, +, X)$. Thus, we can define the set

$$X_0 := \tilde{X} \setminus \bigcup_{i>j} T_i'$$

in $(\mathbb{N}, +, X)$, and for $n \in X_0$, we have $\tilde{F}(n)/V_k(n)$ is bounded above by $k^{p_j}$ and, moreover, this value is achieved on a non-sparse subset of $X_0$.

We now claim that, if $n \in X_0$ and $s \geq 0$, then $k^s n \in X_0$. To see this, suppose that $n \in X_0$ and $k^s n \notin X_0$. Then $n \in X_0 \subseteq \tilde{X}$, so $k^s n \in \tilde{X}$ by Lemma 3.11(iii) and thus $k^s n \in \tilde{X} \setminus X_0$. It follows that $k^s n \in T_i'$ for some $i > j$ and so there is some $\ell$ such that $k^{s+\ell} n \in T_i$. But this gives $n \in T_i'$ by definition and hence $n \notin X_0$, a contradiction.

Since we can define $k^{\mathbb{N}}$ in $(\mathbb{N}, +, X)$ by Fact 2.17, we see we can define the set

$$Y := X_0 \cup k^{\mathbb{N}}$$

and the map $H : Y \to k^{\mathbb{N}}$ given by

$$H(n) = \tilde{F}(n) \cdot k^{-p_j}$$

for $n \in X_0 \setminus k^{\mathbb{N}}$ and $H(n) = n$ for $n \in k^{\mathbb{N}}$. This completes the construction of the set $Y$ and the map $H$, which satisfy properties (i)–(iv). ∎

We can now finally prove our main result of this section. Namely, we show that under the following assumptions on $X \subseteq \mathbb{N}$, the structure $(\mathbb{N}, +, k^{\mathbb{N}})$ defines the function $V_k$. Those assumptions are that $X$ a $k$-automatic, but not definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, and furthermore there is an automaton $\mathcal{A}$ recognizing $X$ such that either $\delta(p, 0) = \{p\}$ for some state $p$ in $\mathcal{A}$, or there exist states $p$ and $p'$ in $\mathcal{A}$ such that $\delta(p, 0) = \varnothing$ or $\delta(p, 0) = \{p'\}$, where $p'$ is in a different strongly connected component from $p$.

**Proof of Theorem 3.1 in Cases I and II** We may assume without loss of generality that $X$ is neither sparse nor eventually periodic. Hence, by Lemma 3.12, we can define

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

a set $Y$ such that properties (i)–(iv) from the statement of the lemma hold. So there is a set $Y_0$ (as described in Lemma 3.12) of $n \in Y$ for which $H(n) = V_k(n)$ contains 1 and is not sparse. Note that this $Y_0$ is not a priori definable in $(\mathbb{N}, +, X)$ since we do not assume that $V_k$ is definable in this structure, though $Y_0$ is clearly $k$-automatic. By [5], there is a natural number $P$ such that every natural number is a sum of at most $P$ elements from $Y_0$.

Now we define a map $G : \mathbb{N} \to k^{\mathbb{N}}$ by the rule

$$G(n) := \max_{r \le P} \max_{\{(i_1,\ldots,i_r) \in Y^r : i_1 + \cdots + i_r = n\}} \min(H(i_1), \ldots, H(i_r)), \tag{17}$$

for $n \ge 1$ and where we take $G(0) = 1$. By construction, this map $G$ is definable in $(\mathbb{N}, +, X)$, since both $Y$ and the map $H$ are.

We claim that $G(n) = V_k(n)$, which will complete the proof of the theorem, since we will then have defined $V_k$ and all $k$-automatic sets are definable from $V_k$. Notice that if $V_k(n) = k^j$ and if $i_1 + \cdots + i_r = n$ with $i_1, \ldots, i_r \in Y$, then we must have $V_k(i_s) \le k^j$ for some $s$. Since $H|_Y \le (V_k)|_Y$, we then see that $H(i_s) \le k^j$ and so $\min(H(i_1), \ldots, H(i_r)) \le k^j$. Thus, we see that $G(n) \le V_k(n)$. We now show that $G(n) \ge V_k(n)$. Write $n = k^s n_0$ with $s \ge 0$, $n_0 \ge 1$, and $k \nmid n_0$. Then we can write $n_0 = i_1 + \cdots + i_r$ for some $r \le P$ and $i_1, \ldots, i_r \in Y_0$. Then $i_1 k^s, \ldots, i_r k^s \in Y_0$ and so by definition of the set $Y_0$, we have $H(i_j k^s) = V_k(i_j k^s) \ge k^s$ for $j = 1, \ldots, r$, and so $G(n) \ge \min(V_k(i_1 k^s), \ldots, V_k(i_r k^s)) \ge k^s$. Thus, we obtain the desired lower bound. This completes the proof for the cases in question. ∎

## 3.2 Proof of Theorem 3.1 in Case (III)

In this section, we give the proof of Theorem 3.1 in Case (III). To handle Case (III), we require a few basic lemmas.

***Lemma 3.13*** *Let $Z$ be a $k$-automatic subset of $\mathbb{N}$ that is not Presburger definable and let $\mathcal{A} = (Q, \{q_0\}, \Sigma_k, \delta, W)$ be a trim, minimized, deterministic automaton that accepts precisely the words that are base-$k$ expansions of elements of $Z$. Then, for each $q \in Q$, the set $Z_q := [L_{q\to q}]_k$ is definable in $(\mathbb{N}, +, Z)$.*

**Proof** Let $E$ denote the language accepted by $\mathcal{A}$. Fix $u \in \Sigma_k^*$ such that $\delta(q_0, u) = q$. Since $\mathcal{A}$ is minimal, the mapping $q \mapsto \chi_q(w)$, defined by

$$\chi_q(w) := \begin{cases} 1 & \text{if } \delta(q, w) \in W, \\ 0 & \text{otherwise,} \end{cases}$$

separates the states. Therefore, there is a finite set of words $w_1, \ldots, w_r \in \Sigma_k^*$ such that each state $q \in Q$ is uniquely determined by the vector $(\chi_q(w_1), \ldots, \chi_q(w_r))$.

Our goal, then, is as follows: given a word $v$, consider what happens when $\mathcal{A}$ is run on $uvw_i$ for each $i$. If $v$ is in $L_{q\to q}$, then the run of $uvw_i$ will begin at $q_0$, proceed to $q$, perform some cycle on the word $v$ that ends back at $q$, and then end at $\delta(q, w_i)$. If $v$ is *not* in $L_{q\to q}$, then $uvw_i$ will instead end at $\delta(q', w_i)$, where $q' = \delta(q, v)$. By our choice of $w_i$, this means that whether $v \in L_{q\to q}$ can be determined by whether each $uvw_i$ is accepted by $\mathcal{A}$, or equivalently, whether each $[uvw_i]_k$ is in $Z$.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

It suffices, therefore, to show that the relation between $[v]_k$ and the tuple $([uvw_i]_k)_i$ is definable in $(\mathbb{N}, +, Z)$, as then, we can define whether $[v]_k \in Z_q$ by deciding whether for each $i$ we have $[uvw_i]_k \in Z$ if and only if $\delta(q, w_i) \in W$. Note that this relation is not a function, because $[\bullet]_k$ is not one-to-one, due to the possibility of leading zeros. We will therefore instead define, for each $i$, the ternary relation of triples $([v]_k, [10^{|v|}]_k, [uvw_i]_k)$, from which the previous relation is definable. This is given by the following formula in $(x, y, z)$:

$$y \in k^{\mathbb{N}} \wedge y > x \wedge z = ([u]_k \cdot y) + (k^{|w_i|} \cdot x) + [w_i]_k.$$

Note that the multiplications by $[u]_k$ and $k^{|w_i|}$ are constant multiplications and hence Presburger-definable. The only remaining matter is how we mean to use $y \in k^{\mathbb{N}}$ in the above formula. But because $Z$ is $k$-automatic and not Presburger-definable, $k^{\mathbb{N}}$ is definable in $(\mathbb{N}, +, Z)$ by Fact 2.17. ∎

***Lemma 3.14*** *Let $\mathcal{A} = (Q, \{q_0\}, \Sigma_k, \delta, W)$ be a trim, minimized, deterministic automaton and let $q, q', q'' \in Q$ be in same strongly connected component. Then we have:*

(a) *If $[L_{q' \to q'}]_k$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, then so is $[L_{q \to q''}]_k$.*
(b) *If $[L_{q' \to q'}]_k$ is sparse, then so is $[L_{q \to q''}]_k$.*

**Proof** We will once again use Fact 2.17 to work in the expanded language $\{+, <, k^{\mathbb{N}}, \ldots\}$. To prove (a), since $q'$ and $q$ are in the same strongly connected component, there is some word $u \in L_{q' \to q}$ and so if $w$ is a word with no leading zeros, then $[w]_k \in [L_{q \to q'}]_k$ if and only if $0^i w \in L_{q \to q'}$ for some $i$. By the pumping lemma, there is some pumping length $p$ such that we can take $i < p$. Then

$$[w]_k \in [L_{q \to q'}]_k \iff \exists i < p,\ \ 0^i w \in L_{q \to q'} \iff \exists i < p,\ \ u0^i w \in L_{q' \to q'}.$$

Then $n = [w]_k \in [L_{q \to q'}]_k$ if and only if

$$\exists s \in k^{\mathbb{N}}, (n < s \leq (k^p \cdot n)) \wedge (([u]_k \cdot s) + n \in [L_{q' \to q'}]_k)$$

(here $k^p$ and $[u]_k$ are constants). Since $[L_{q' \to q'}]_k$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, we then see that $[L_{q \to q'}]_k$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$.

Similarly, let $w'$ be a word such that $\delta(q'', w') = q'$, which is guaranteed to exist because $q', q''$ are in the same strongly connected component. Let $v$ be a word without leading zeros that may or may not be in $L_{q \to q''}$, and let $n = [v]_k$. Then:

$$n \in [L_{q \to q''}]_k \iff vw' \in L_{q \to q'} \iff k^{|w'|} n + [w']_k \in [L_{q \to q'}]_k.$$

So $L_{q \to q''}$ is also definable in $(\mathbb{N}, +, k^{\mathbb{N}})$.

To prove (b), observe that by [6, Proposition 7.1(3)], if $L_{q \to q'}$ is not sparse, then there exist words $u, v, a, b$ with $a, b$ distinct and of the same length such that $L_{q \to q'} \supseteq u\{a, b\}^* v$. But now since $q, q'$ are in the same strongly connected component, there exists $w \in L_{q' \to q}$ and $w' \in L_{q' \to q''}$ and so $L_{q' \to q'} \supseteq wu\{a, b\}^* vw'$, which, again using [6, Proposition 7.1(3)], contradicts that $L_{q' \to q'}$ is sparse. ∎

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

We are finally able to prove Theorem 3.1 in Case (III). As a reminder, this is the case where we assume that 0 is idempotent with respect to $\mathcal{A}$ and that $\delta(p, 0)$ contains a distinct state $p'$ in the same strongly connected component as $p$.

**Proof of Theorem 3.1 in Case (III)** We let $p' \in Q$ be such that $\delta(p, 0) = \{p'\}$. Since 0 is idempotent, we have $\delta(p', 0) = \{p'\}$ and, by assumption, $p$ and $p'$ are in the same strongly connected component. Since $\delta(p', 0) = \{p'\}$, we see that the cycle language $L_{p' \to p'}$ falls into Case I, and hence is either definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, or it defines $V_k$. If expanding $(\mathbb{N}, +)$ by a predicate for $[L_{p' \to p'}]_k$ generates the same definable sets as $(\mathbb{N}, +, k^{\mathbb{N}})$, then by taking the subautomaton induced by this strongly connected component containing $p'$ (but where we take our initial state to be $p$ and our final state to be $p$) we see by Lemma 3.13 that $[L_{p \to p}]_k$ defines $[L_{p' \to p'}]_k$ over $(\mathbb{N}, +)$, and hence also defines $V_k$. Thus, we may assume that $[L_{p' \to p'}]$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, and hence, by Lemma 3.14, so is $[L_{p \to p}]_k$. Thus, we have proven the dichotomy in this case. ∎

## 4 Proof of dichotomy: General case

We now show how to prove Theorem 1.1 from Theorem 3.1. More precisely, we prove that if $X$ is a $k$-automatic set that is not definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, then $X$ defines $V_k$.

We need two lemmas before giving the proof of the general dichotomy.

***Lemma 4.1*** *Let $\mathcal{A} = (Q, \{q_0\}, \Sigma_k, \delta, F)$ be a deterministic, trim, strongly connected automaton. If there exists a state $q \in Q$ and a character $c \in \Sigma_k$ such that $\delta(q, c) = \varnothing$, then there exists a "forbidden subword" for $\mathcal{A}$, that is, a word $w$ such that no word accepted by $\mathcal{A}$ contains $w$ as a subword.*

**Proof** We will construct $w$ iteratively. Begin by letting $w_0 = \varepsilon$. Enumerate the states in $Q$ as $Q = \{q_1, \ldots, q_n\}$.

Our goal is to construct each $w_i$ such that $\delta(q_j, w_i) = \varnothing$ for each $1 \leq j \leq i$. So to construct $w_i$ from $w_{i-1}$, consider $\delta(q_i, w_{i-1})$, that is, determine which state $p_i$, if any, results from running the word $w_{i-1}$ starting from $q_i$. (By determinism, there is at most one such state.) If there is no such state, we may let $w_i = w_{i-1}$. Otherwise, note that by strong connectedness and determinism of $\mathcal{A}$, there is a word $v_i$ such that $\delta(p_i, v_i) = \{q\}$.

We then let $w_i = w_{i-1} v_i c$. Note that:

$$\begin{aligned} \delta(q_i, w_i) &= \delta(\delta(\delta(q_i, w_{i-1}), v_i), c) \\ &= \delta(\delta(p_i, v_i), c) \\ &= \delta(q, c) \\ &= \varnothing. \end{aligned}$$

Moreover, note that suffixing words onto $w_{i-1}$ does not change the fact that $\delta(q_j, w_i) = \delta(q_j, w_{i-1}) = \varnothing$ for $j < i$. In other words, we achieve the desired effect that $\delta(q_j, w_i) = \varnothing$ for each $1 \leq j \leq i$.

We claim that no word accepted by $\mathcal{A}$ contains $w_n$ as a subword. Consider any word of the form $u w_n v$, and let $i$ be such that $\delta(q_0, u) = q_i$. Then when $\mathcal{A}$ is run on this word:

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

$$\begin{aligned}\delta(q_0, uw_n v) &= \delta(\delta(\delta(q_0, u), w_n), v)\\ &= \delta(\delta(q_i, w_n), v)\\ &= \delta(\varnothing, v)\\ &= \varnothing.\end{aligned}$$

So the word $uw_n v$ is rejected by $\mathcal{A}$. ∎

For the next lemma, recall that in the context of Semenov's characterization defined in Section 2.4, $\Sigma_{\ell,m,c}$ is the set of all words of length a multiple of $\ell$ that are the base-$k$ expansion (*without* leading zeros) of a word congruent to $-c$ modulo $m$.

***Lemma 4.2*** *Let $\ell, m, c$ be integers with $m, \ell > 0$, and let $u \in \Sigma_k^*$. Then the language $L = u\Sigma_{\ell,m,c}$ has no "forbidden subword," that is, every $w \in \Sigma_k^*$ is a subword of some word in $L$.*

**Proof** Given $w \in \Sigma_k^*$, we will find an element of $L$ containing $w$ as a subword. Pick $M$ such that $\ell M > |w|$ and such that $(k-1)k^{\ell M - |w| - 1} \geq m + 1$. Now consider the language $M = uw\Sigma_k^{\ell M - |w|}$. Every word in $M$ has $w$ as a subword, so it will suffice to show that $L \cap M$ is nonempty.

By definition of $L$ and $M$, $L \cap M$ will contain precisely those words $uwv$ such that $|v| = \ell M - |w|$ and $wv \in \Sigma_{\ell,m,c}$. Note that the length condition of $\Sigma_{\ell,m,c}$ is satisfied, because $|wv| = |w| + |v| = |w| + \ell M - |w| = \ell M$ is a multiple of $\ell$. The remaining condition on $v$ is that $wv$ must be the base-$k$ expansion without leading zeros of a word congruent to $-c$ modulo $m$. In other words, $L \cap M$ will be nonempty as long as there is a word $v$ of length $\ell M - |w|$ beginning with a nonzero character and such that $[wv]_k \equiv -c \pmod{m}$.

Consider the set:

$$S = \{[wv]_k : v \in \Sigma_k^{\ell M - |w|}, v \text{ does not begin with } 0\},$$

and observe that we may write:

$$S = \mathbb{Z} \cap \left[\left[w10^{\ell M - |w| - 1}\right]_k, \left[w(k-1)^{\ell M - |w|}\right]_k\right),$$

that is, $S$ is a set of consecutive integers. By a counting argument, $|S| = (k-1)k^{\ell M - |w| - 1} \geq m + 1$. Because $S$ is a set of more than $m$ consecutive integers, it follows that some $s \in S$ is congruent to $-c$ modulo $m$. So by letting $v$ be such that $[wv]_k = s$, we obtain a word $uwv \in L \cap M$. ∎

We now prove Theorem 1.1.

**Proof of Theorem 1.1** Assume toward a contradiction that there is some $k$-automatic set $X$ that is not definable in $(\mathbb{N}, +, k^{\mathbb{N}})$ but which does not define $V_k$, and pick a minimal DFA $\mathcal{A}$ that accepts the language $L$ consisting of words whose base-$k$ expansion lies in $X$. By Lemma 3.13, we have that $X$ defines $X_q := [L_{q \to q}]_k$ for all $q \in Q$. Hence, by Theorem 3.1, since $X$ does not define $V_k$, we have that $X_q$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$ for all $q \in Q$.

We let $C_1, \ldots, C_s$ denote the strongly connected components of $\mathcal{A}$. Recall that some $C_i$ are leaves, which, by Definition 2.11, are strongly connected components containing a transition to a different strongly connected component. We claim that if $q \in C_i$ for such a component, $X_q$ must be sparse.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

To this end, fix such an $i$ and $q \in C_i$, and let $p \in C_i, y \in \Sigma_k$ be such that $\delta(p, y)$ lies outside $C_i$. Then $X_q$ is recognized by the automaton formed by taking $C_i$ as the set of states, modifying $\delta$ to a new transition function $\delta'$ containing only transitions within $C_i$, and making $q$ the only initial and final state; in particular, note that $\delta'(p, y) = \varnothing$.

By Lemma 4.1, there is a word $w$ that is not a subword of any word accepted by this new automaton, and therefore by the contrapositive of Lemma 4.2, the language of $X_q$ contains no subset of the form $u\Sigma_{\ell,m,c}$. We saw earlier that $X_q$ was definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, so by Fact 2.19, $X_q$ is a union of sparse sets and is thus sparse. Furthermore, by Lemma 3.14, $L_{q\to q'}$ must be sparse for all $q, q' \in C_i$.

Now, to each word $w$ in $L$, we associate a finite sequence of states $(p_1, q_1, p_2, q_2, \ldots, p_n, q_n)$ and a finite sequence of characters $(\sigma_1, \ldots, \sigma_{n-1})$ such that the run of $w$ in $\mathcal{A}$:

- begins in the initial state $p_1$;
- proceeds to $q_1$, in the same strongly connected component as $p_1$;
- transitions on the character $\sigma_1$ to $p_2$, in a different strongly connected component;
- proceeds to $q_2$, in the same strongly connected component as $p_2$;

  ⋮
- finishes in $q_n$, an accept state.

There are only finitely many choices of the finite sequences $(p_1, q_1, p_2, q_2, \ldots, p_n, q_n)$ and $(\sigma_1, \ldots, \sigma_{n-1})$, so we see that $L$ is a finite union of languages of the form

$$M = L_{p_1\to q_1}\sigma_1 L_{p_2\to q_2}\sigma_2 \ldots L_{p_{n-1}\to q_{n-1}}\sigma_{n-1}L_{p_n\to q_n},$$

where:

- $p_i$ and $q_i$ are in the same strongly connected component;
- $\sigma_i \in \Sigma_k$ and $\delta(q_i, \sigma_i) = p_{i+1}$ for $i = 1, \ldots, n-1$;
- $q_i, p_{i+1}$ are in different strongly connected components.

We therefore aim to achieve a contradiction by proving that for each such language $M$, the corresponding set $[M]_k$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$. This will show that $X$ itself is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, directly contradicting a premise.

Consider one such $M = L_{p_1\to q_1}\sigma_1 L_{p_2\to q_2}\sigma_2 \ldots L_{p_{n-1}\to q_{n-1}}\sigma_{n-1}L_{p_n\to q_n}$, and note that each $p_i$ is in the same strongly connected component as the corresponding $q_i$. Recall that each $X_{p_i}$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$; by Lemma 3.14, so is $L_{p_i\to q_i}$. Moreover, note that $L_{p_1\to q_1}, \ldots, L_{p_{n-1}\to q_{n-1}}$ are path languages within strongly connected components of $\mathcal{A}$ that are not leaves, in which case we showed earlier that each of them is sparse.

Therefore, the language $M' = L_{p_1\to q_1}\sigma_1 L_{p_2\to q_2}\sigma_2 \ldots L_{p_{n-1}\to q_{n-1}}\sigma_{n-1}$ is sparse, because it is a concatenation of sparse languages (cf. [6, Proposition 7.1(6)]). So $M = M'L_{p_n\to q_n}$ is the concatenation of a sparse language and a language whose corresponding subset $[L_{p_n\to q_n}]_k \subseteq \mathbb{N}$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$. By Fact 2.19, $[M]_k = [M'L_{p_n\to q_n}]_k$ is also definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, and so as stated above, we achieve our contradiction, because $X$ is shown to be a finite union of sets definable in $(\mathbb{N}, +, k^{\mathbb{N}})$. ∎

## 5 Further results

The characterization established in the previous section has consequences in the realms of both definability and decidability. Bès previously established in [7] that for

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

$k$ and $\ell$ multiplicatively independent natural numbers, the structure $(\mathbb{N}, +, V_k, \ell^{\mathbb{N}})$ defines multiplication and hence is undecidable. Using a theorem of the third-named author showing that under the same hypotheses $(\mathbb{N}, +, k^{\mathbb{N}}, \ell^{\mathbb{N}})$ does not define multiplication [17], we extend that result with the following corollary.

***Corollary 5.1*** *If $k, \ell \in \mathbb{N}_{>1}$ are multiplicatively independent, and $X \subseteq \mathbb{N}$ is $k$-automatic and $Y \subseteq \mathbb{N}$ is $\ell$-automatic, then the structure $(\mathbb{N}, +, X, Y)$ defines multiplication if and only if either $X$ is not definable in $(\mathbb{N}, +, k^{\mathbb{N}})$ or $Y$ is not definable in $(\mathbb{N}, +, \ell^{\mathbb{N}})$.*

Using results of Hawthorne [14] relating definability in $(\mathbb{Z}, +, k^{\mathbb{N}})$ to stability of the structure $(\mathbb{Z}, +, X)$, where $X \subseteq \mathbb{N}$ is $k$-automatic, we also obtain a further characterization of definability of $V_k$ in such expansions of $(\mathbb{Z}, +)$.

***Corollary 5.2*** *Suppose that $X \subseteq \mathbb{N}$ is $k$-automatic. Then exactly one of the following holds:*

(1) *The structure $(\mathbb{Z}, +, X)$ is stable.*
(2) *The structures $(\mathbb{Z}, +, X)$ and $(\mathbb{Z}, <, +, k^{\mathbb{N}})$ define the same sets.*
(3) *The structure $(\mathbb{Z}, +, X)$ defines $V_k$.*

**Proof** First, we note that in [14], Hawthorne shows that if $X \subseteq \mathbb{N}$ is $k$-automatic, the structure $(\mathbb{Z}, +, X)$ is stable if and only if $X$ is sparse. In the case that $X$ is not sparse, by results of Bell et al. in [5], there exist natural numbers $m, d, N$ such that every natural number $n > N$ that is a multiple of $d$ can be written as the sum of at most $m$ elements of $X$. From this, we can define $\mathbb{N}$ in $(\mathbb{Z}, +, X)$. Having defined $\mathbb{N}$, we may now apply Theorem 1.1 to the induced structure on $\mathbb{N}$. Hence, we conclude that either $(\mathbb{Z}, +, X)$ defines the same sets as $(\mathbb{Z}, <, +, k^{\mathbb{N}})$ in the case that $\mathbb{N} \backslash X$ is sparse, or else $(\mathbb{Z}, +, X)$ defines $V_k$. ∎

This corollary prompts us to ask the following question.

***Question 5.3*** *If $X \subseteq \mathbb{Z}$ is recognized by an automaton with alphabet $\Sigma = \{-, 0, \ldots, k-1\}$, does the above trichotomy hold?*

We note that for $X \subseteq \mathbb{Z}^2$, this is not the case. For $k \in \mathbb{N}_{>1}$, consider $<_k := \{(x, y) \in \mathbb{Z}^2 : V_k(|x|) < V_k(|y|)\}$. By results of [3], the structure $(\mathbb{Z}, +, <_k)$ is dp-minimal for all $k > 1$, so $V_k$ cannot be defined in this structure. However, the classic ordering $<$ of the integers is also not definable in this structure, yet the structure is nevertheless unstable.

It is also natural to ask whether Theorem 1.1 holds when we consider a $k$-automatic set $X \subseteq \mathbb{N}^m$ with $m > 1$. One of the obstacles in proving the higher-arity analog is the fact that Semenov only gives a nice characterization of one-dimensional definable sets in [18]. This prompts the following question.

***Question 5.4*** *Suppose $X \subseteq \mathbb{N}^m$, with $m > 1$, is $k$-automatic. If the unary sets definable in $(\mathbb{N}, +, X)$ are precisely the unary sets in $(\mathbb{N}, +, k^{\mathbb{N}})$, is $X$ necessarily definable in $(\mathbb{N}, +, k^{\mathbb{N}})$?*

As illustrated by the structure $(\mathbb{Z}, +, <_k)$, this question has a negative answer when we replace $X \subseteq \mathbb{N}^m$ with $X \subseteq \mathbb{Z}^m$ (for $m > 1$).

It is also quite natural to ask whether we can effectively decide whether a given $k$-automatic set $X \subseteq \mathbb{N}$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$. By Semenov's characterization of

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.

definable unary sets in $(\mathbb{N}, +, k^{\mathbb{N}})$, we know that for any $X \subseteq \mathbb{N}$ definable in this structure, there is a set $Y$ definable in Presburger arithmetic such that the symmetric difference of $X$ and $Y$ is sparse. Hence, this decidability question boils down to deciding whether there exists some $N \in \mathbb{N}$ after which $\{x \in X : x > N\}$ is periodic, as demonstrated by our results. Certainly, we can write down an algorithm that enumerates all $X$ definable in $(\mathbb{N}, +, k^{\mathbb{N}})$, but deciding whether $X$ is definable in $(\mathbb{N}, +, k^{\mathbb{N}})$ for an arbitrary $X$ definable in $(\mathbb{N}, +, V_k)$ seems strictly trickier.

## References


[1] J.-P. Allouche and J. Shallit, *The ring of k-regular sequences*. Theor. Comput. Sci. **98**(1992), 163–197.

[2] J.-P. Allouche and J. Shallit, *Automatic sequences: Theory, applications, generalizations*, Cambridge University Press, Cambridge, 2003.

[3] E. Alouf and C. d'Elbée, *A new dp-minimal expansion of the integers*. J. Symbolic Logic **84**(2019), no. 2, 632–663.

[4] P.-G. Becker, *k-regular power series and Mahler-type functional equations*. J. Number Theory **49**(1994), 269–286.

[5] J. P. Bell, K. Hare, and J. Shallit, *When is an automatic set an additive basis?* Proc. Amer. Math. Soc. Ser. B **5**(2018), 50–63.

[6] J. P. Bell and R. Moosa, *F-sets and finite automata*. J. Théor. Nombres Bordeaux **31**(2019), no. 1, 101–130.

[7] A. Bès, *Undecidable extensions of Büchi arithmetic and Cobham-Semenov theorem*. J. Symbolic Logic **62**(1997), no. 4, 1280–1296.

[8] V. Bruyère, *Entiers et automates finis*. Mémoir de fin d'études, 1985.

[9] J. R. Büchi, *Weak second-order arithmetic and finite automata*. Z. Math. Logik Grundlagen Math. **6**(1960), 66–92.

[10] J. R. Büchi, *On a decision method in restricted second order arithmetic*. In: E. Nagel, P. Suppes, A. Tarski (eds.), *Logic, methodology and philosophy of science (Proceedings of the 1960 International Congress)*, Stanford University Press, Stanford, 1962, pp. 1–11.

[11] A. Cobham, *On the base-dependence of sets of numbers recognizable by finite automata*. Math. Syst. Theory **3**(1969), no. 2, 186–192.

[12] H. B. Enderton, *A mathematical introduction to logic*, Harcourt/Academic Press, University of Michigan, Ann Arbor, 2001.

[13] S. Ginsburg and E. H. Spanier, *Semigroups, Presburger formulas, and languages*. Pacific J. Math. **16**(1966), no. 2, 285–296.

[14] C. Hawthorne, *Automata and tame expansions of* $(\mathbb{Z}, +)$. Israel J. Math. **249**(2022), 651–693.

[15] Q. Lambotte and F. Point, *On expansions of* $(\mathbb{Z}, +, 0)$. Ann. Pure Appl. Logic **171**(2020), no. 8, 102–809.

[16] M. 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 des Mathématiciens des Pays Slaves*, Warszawa, 1929, pp. 92–101.

[17] C. Schulz, *Undefinability of multiplication in Presburger arithmetic with sets of powers*. J. Symbolic Logic **90**(2025), no. 2, 872–886.

[18] A. L. Semenov, *On certain extensions of the arithmetic of addition of natural numbers*. Izv. Math. **15**(1980), no. 2, 401–418.

[19] R. Villemaire, *The theory of* $(\mathbb{N}, +, V_k, V_\ell)$ *is undecidable*. Theor. Comput. Sci. **106**(1992), no. 2, 337–349.



*University of Waterloo, Canada*
*e-mail*: jpbell@uwaterloo.ca chris.schulz@uwaterloo.ca

*Universiteit van Amsterdam, Netherlands*
*e-mail*: a.t.blockgorman@uva.nl


Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 13:52:38, subject to the Cambridge Core terms of use.