Algebraic Characterizations and Block Product Decompositions for First Order Logic and its Infinitary Quantifier Extensions over Countable Words

Bharat Adsul[a], Saptarshi Sarkar[a], A. V. Sreejith[b,*]

[a]*Indian Institute of Technology Bombay, Powai, Mumbai, 400076, Maharashtra, India*
[b]*Indian Institute of Technology Goa, Farmagudi, Ponda, 403401, Goa, India*

Abstract

We contribute to the refined understanding of language-logic-algebra interplay in a recent algebraic framework over countable words. Algebraic characterizations of the one variable fragment of FO as well as the boolean closure of the existential fragment of FO are established. We develop a seamless integration of the block product operation in the countable setting, and generalize well-known decompositional characterizations of FO and its two variable fragment. We propose an extension of FO admitting infinitary quantifiers to reason about inherent infinitary properties of countable words, and obtain a natural hierarchical block-product based characterization of this extension. Properties expressible in this extension can be simultaneously expressed in the classical logical systems such as WMSO and FO[cut]. We also rule out the possibility of a finite-basis for a block-product based characterization of these logical systems. Finally, we report algebraic characterizations of one variable fragments of the hierarchies of the new extension.

Keywords: linear orderings, first-order logic, countable words, algebraic structures, formal language theory, block product, Krohn-Rhodes theorem

*Corresponding author

Email addresses: `adsul@cse.iitb.ac.in` (Bharat Adsul), `sreejithav@iitgoa.ac.in` (A. V. Sreejith)

URL: `https://www.iitgoa.ac.in/~sreejithav/` (A. V. Sreejith)

1. Introduction

Monadic Second-Order (MSO) logic is a natural logic to express properties of words. Over finite words, Büchi-Elgot-Trakhtenbrot theorem [1] establishes that languages definable in MSO are precisely *regular* languages. Regular languages admit a variety of well-known characterizations [1, 2, 3] such as describability by regular expressions, acceptance by finite state automata, or recognition by finite monoids. The seminal results of Büchi [4], Rabin [5], Shelah [6], and Carton et.al [7] show that this close relationship between logical expressiveness and language recognizability remains true not just over finite linear orderings but also over infinite words like ω-words and countable words. The effective translation between MSO and an automata/algebra model gives decidability of MSO over these linear orderings. The classical result of Shelah (also in [6]) shows that over reals (uncountable orderings) MSO is undecidable. In this paper, we focus on analysing the expressive power and decidability of various logics over countable words.

One can effectively associate, to a regular language of finite words, its *syntactic monoid*. This canonical algebraic structure carries a rich amount of information about the corresponding language. Its role is highlighted by the classical Schützenberger-McNaughton-Papert theorem [8, 9] which shows that *aperiodicity* property of the syntactic monoid coincides with describability using star-free expressions as well as definability in First-Order (FO) logic. Building on the work of Shelah [6], Carton et. al. [7] proposed an algebraic model, $\circledast$-monoid, that recognize exactly those languages definable by MSO over countable linear orderings. This framework extends the language-logic-algebra interplay to the setting of countable words. The algebraic approach paves the way for equational characterizations of logics and hence decidability of the problem of definability in the said logics. Building on the work in [7], algebraic characterization for variety of sub-logics of MSO over countable words is carried out in [10]. In particular, this work provides algebraic *equational* (hence decidable) characterizations of $\mathrm{FO}, \mathrm{FO}[\mathrm{cut}]$ – an *extension* of FO that allows quantification over *Dedekind-cuts* and WMSO – an *extension* of FO that allows quantification over finite sets. A decidable algebraic equational characterization for the two variable fragment of FO (denoted by FO^2) over countable words is presented in [11].

We begin our explorations in Section 3 with the *small* fragments of FO over countable words. We provide an equational characterization (Theorem 3) for FO^1 – the one variable fragment of FO. Coupled with the results in [11] and [10] on the equational characterization of FO^2 and $FO = FO^3$ (see [12]), we have complete *equational* characterizations of FO fragments defined by the number of permissible variables. Our next result in the same section (Theorem 4) extends Simon's theorem on piecewise testable languages to countable words and provides a natural algebraic characterization of the Boolean closure of the existential-fragment of FO.

It turns out that the algebraic landscape of small fragments of FO over countable words parallels very closely the same landscape over finite words. This can be attributed to the limited expressive power of FO over countable words. For instance, Bès and Carton [13] showed that the seemingly natural 'finiteness' property (that the set of all positions is a finite set) of countable words can not be expressed in FO!

In Section 6 we extend FO with new *infinitary* quantifiers. The main purpose of our new quantifiers is to naturally allow expression of infinitary features that are inherent in the countable setting. An example formula using such an infinitary quantifier is: $\exists^{\infty_1} x\ a(x) \wedge \neg\exists^{\infty_1} x\ b(x)$. In its natural semantics, this formula with one variable asserts that there are infinitely many a-labelled positions and only finitely many b-labelled positions. We propose an extension of FO called $FO[\infty]$ that supports *first-order infinitary quantifiers* of the form $\exists^{\infty_k} x$ to talk about existence of higher-level infinitely (more accurately, "Infinitary rank" k) many witnesses x. We organize $FO[\infty]$ in a natural hierarchy based on the maximum allowed infinitary-level of the quantifiers. We prove that $FO[\infty]$ properties can be expressed simultaneously (Theorem 8) in FO[cut] as well as WMSO.

The other main results of this work are decomposition theorems in the countable setting. The seminal result of Krohn-Rhodes decomposition theorem [14] shows that any finite monoid can be built from groups and the monoid U_1 (a unique 2-element monoid) using a block-product construction [15]. There are other prominent examples in this line of work. A characterization of FO-logic (resp. FO^2, the two-variable fragment) in terms of strongly (resp. weakly) iterated block-products of copies of U_1 is presented in [15] (resp. [16]).

Motivated by the decisive role played by block products in the standard settings [15, 3], we introduce block products in the countable setting in Section 4. The block product construction associates to a pair of $\circledast$-monoids (more precisely, $\oplus$-semigroups) (M, N) a new $\circledast$-monoid (more precisely, $\oplus$-semigroup) $M \Box N$. From a formal-language theoretic viewpoint, the importance of the block product construction is best described by the accompanying block product principle (Theorem 5). Roughly speaking the block product principle says that evaluating a *countable* word u in $M \Box N$ can be achieved by the following two-stage process:

1. evaluate the word u *in* M and label every position x of u with the additional information about evaluations of $u_{<x}$ and $u_{>x}$ *in* M where $u_{<x}$ and $u_{>x}$ are such that $u = u_{<x}u[x]u_{>x}$ (that is, $u_{<x}$ and $u_{>x}$ are the left and right factors/contexts at position x);

2. evaluate this extended word (with the additional information) *in* N.

Said differently, M 'operates' on u as usual; while when N 'operates' on u, it has access, at *every* position, to evaluations of M on left-right contexts at that position. Our block product construction and the accompanying block product principle extend naturally the results from finite words to countable words. Furthermore, we give decompositional characterizations of FO and FO^2 over countable words (Theorems 7 and 6 respectively) - again natural extensions of analogous results over finite words.

In Section 7, we extend the block-product based characterization of FO to FO[∞] (Theorem 10). Towards this, we identify an appropriate simple family of $\circledast$-algebra and show that this family (in fact, its initial fragments) serve as a basis in our hierarchical block-product based characterization. We also show that the language-logic-algebra connection for FO^1 admits novel generalizations to the one variable fragments of the new hierarchical extensions.

In Section 8, we present a 'no finite block-product basis' theorem (Theorem 12) for FO[∞], FO[cut], and the semantic class FO[cut] $\cap$ WMSO. The theorem states that no finite set of $\circledast$-algebras closed under block products recognize all languages definable in these logics. This is in contrast with FO where the unique 2-element $\circledast$-algebra is a basis for a block-product based characterization. To prove the above result we identify a natural combinatorial measure called *gap-nesting-length* that is shown to be well-behaved with respect to the block product operation.

The rest of the article is organized as follows. Section 2 recalls basic notions about countable words and summarizes the necessary algebraic background from the framework [7]. Section 3 deals with the small fragments of FO: FO^1 and the Boolean closure of the existential fragment of FO. In Section 4 and Section 5 we develop the algebraic apparatus of block product operation and weakly iterated block-product based characterization of FO^2.

Section 6 is devoted to FO[∞] and its relation with FO[cut] and WMSO and in Section 7, we provide the relevant characterizations. Section 8 is concerned with the 'no finite block-product basis' theorems. We finally conclude in Section 9.

The results presented in Sections 3, 6, 7, and 8 are an elaboration and extension of the work that appeared in FCT 2021 [17]. In order to make this article self-contained, we have also included relevant work of the authors (Sections 4, and 5) that was presented in LICS 2019 [18]. This paper includes the full proofs of the results, many of which are not present in the conference proceedings.

2. Preliminaries

In this section, we briefly present some mathematical preliminaries of countable linear orderings, and recall the algebraic framework developed in [7].

A *countable linear ordering* (or simply ordering) $\alpha = (X, <)$ is a countable set X equipped with a total order $<$. An ordering $\beta = (Y, <)$ is called a *subordering* of α if $Y \subseteq X$ and the order on Y is induced from that on X. We denote by ω, ω^* and η the orderings $(\mathbb{N}, <), (-\mathbb{N}, <)$ and $(\mathbb{Q}, <)$ respectively. A subordering $(I, <)$ of α is called *convex* if for any $x < y \in I$, and $z \in \alpha$, $x < z < y$ implies $z \in I$. A subordering $(I, <)$ of α is *dense* in α if for any two points $x < y \in \alpha$, there exists $z \in I$ such that $x < z < y$. For example, $(\mathbb{Q}, <)$ is dense in $(\mathbb{R}, <)$ and $(\mathbb{R}, <)$ is dense in itself. If a linear ordering is dense in itself, we simply call it dense. A linear ordering is called *scattered* if all its dense suborderings are singleton or empty. The *generalized sum* of countably many (disjoint) linear orderings $\beta_i = (X_i, <_i)$ which are themselves indexed by some linear ordering $\alpha = (Y, <)$ is the linear ordering $\sum_{i \in \alpha} \beta_i = (Z, <')$ where $Z = \bigcup_{i \in \alpha} X_i$ and for any two points $x, y \in Z$, $x <' y$ if either $x <_i y$ or $x \in X_i$, $y \in X_j$ and $i < j$. The book [19] contains an in-depth study of linear orderings.

A countable word w is a labelled countable linear ordering. More formally, given a finite alphabet Σ and a countable linear ordering α, a countable word (or simply word) w is a map $w : \alpha \to \Sigma$. We call α the *domain* of w, denoted $\mathrm{dom}(w)$. For a word w, we say a point or position x in the word to refer to an element of its domain. The notation $w[x]$ denotes the letter at the x^{th} position in the word w. A *subword* is a restriction of a word w to some induced subordering I of its domain, and is denoted by w_I. If I is convex, then w_I is called a *factor*.

For two countable words u and v, we will denote by uv the countable word formed by the concatenation of u and v. The *generalized concatenation* of a countable sequence of words $(u_i)_{i\in\alpha}$ indexed by a linear countable ordering α is the unique word $\prod_{i\in\alpha} u_i = v$ where $\mathrm{dom}(v) = \sum_{i\in\alpha} \mathrm{dom}(u_i)$, and $v[x] = u_i[x]$ if $x \in \mathrm{dom}(u_i)$.

The following countable words are of special interest. The notation ε stands for the *empty word* (the word over the empty domain). The *ω-word*, a^ω denotes the word over the domain $(\mathbb{N}, <)$ such that every position is mapped to the letter a. Similarly, the *ω^*-word* a^{ω^*} denotes the word over the domain $(-\mathbb{N}, <)$ where every position is mapped to letter a. A *perfect shuffle* over a nonempty set $P \subseteq \Sigma$ of letters, denoted by P^η, is the word w over domain $(\mathbb{Q}, <)$ such that $w[x] \in P$ for all positions x in $\mathrm{dom}(w)$ and for any $a \in P$, any $x < y$ in $\mathrm{dom}(w)$, there exists $z \in \mathrm{dom}(w)$ such that $w[z] = a$ and $x < z < y$. This is a unique word up to isomorphism [19].

Example 1. The word $(a^\omega)^\omega$ denotes the countable word formed by generalized concatenation of ω many words a^ω. Similarly, for any countable word u, the word u^{ω^*} denotes the countable word formed by generalized concatenation of ω^* many words u. Note that upto isomorphism the words $(a^\eta)^\omega$, $(a^\eta)^{\omega^*}$, and $(a^\eta)^\eta$, is the same word.

For an alphabet Σ, the set of all countable words is denoted by $\Sigma^{\circledast}$ and the set of all countable words over non-empty domain is denoted by $\Sigma^{\oplus}$.

We now recall the algebraic framework from [7]. A *$\oplus$-semigroup* (S, π) consists of a set S with an operation $\pi : S^{\oplus} \to S$ such that, $\pi(a) = a$ for all $a \in S$ and π satisfies the *generalized associativity property* – that is $\pi\big(\prod_{i\in\alpha} u_i\big) = \pi\big(\prod_{i\in\alpha} \pi(u_i)\big)$ for every countable linear ordering α. If the generalized associativity holds with $\pi : S^{\circledast} \to S$, then (S, π) is called a $\circledast$-monoid. It is easy to see that, in this case, the element $1 = \pi(\varepsilon)$ of S is the *neutral* element of S. The defining property of a neutral element 1 is that: for every word $u \in S^{\oplus}$, if the word $u|_{\neq 1}$ is the subword obtained by removing every occurence of the element 1 and $u|_{\neq 1}$ is non-empty, then $\pi(u) = \pi(u|_{\neq 1})$.

It is easy to see that if a given $\oplus$-semigroup (S, π) does not admit a neutral element, we can construct a $\circledast$-monoid on the set $S^1 = S \cup 1$ by introducing an *additional* element 1 and by extending π suitably to $S^{1^{\circledast}}$ so that 1 becomes the neutral element. On the other hand, if $\oplus$-semigroup contains a neutral element, say $1 \in S$, then (S, π) is already a $\circledast$-monoid with $\pi(\varepsilon) = 1$. In this case, we simply set $S^1 = S$.

A $\oplus$-semigroup or $\circledast$-monoid (S, π) is called finite if S is finite. For a set Σ, $(\Sigma^{\oplus}, \prod)$ (resp. $(\Sigma^{\circledast}, \prod)$) is the *free* $\oplus$-semigroup (resp. free $\circledast$-monoid) generated by Σ.

Example 2. $\mathrm{U}_1 = (\{1, 0\}, \pi)$ is a finite $\circledast$-monoid where π is defined for all $u \in \{1, 0\}^{\circledast}$ as:

$$\pi(u) = \begin{cases} 1 & \text{if } u \in \{1\}^{\circledast} \\ 0 & \text{otherwise} \end{cases}$$

Here π satisfies the generalized associativity property because no matter which factorization of u we take, the output of π applied directly on u equals the output of π applied in two stages — first on the factors, and then on the countable word formed by the outputs of the previous stage. Let us consider the word $u = (011)^{\omega}$. We have $\pi(u) = 0$ since u contains 0. If we consider the factorization $u = \prod_{i \in \omega}(011)$, then note that $\pi(\prod_{i \in \omega}(\pi(011)) = \pi(\prod_{i \in \omega} 0) = 0$ which indeed equals $\pi(u)$.

Let (S, π) be a $\oplus$-semigroup. Note that even if S is finite, π need not be finitely presentable and hence not suitable for algorithmic purposes. Fortunately, it is possible to capture π through finitely presentable operators. This is precisely the reason for introducing $\oplus$-algebras.

A $\oplus$-*algebra* $(S, \cdot, \tau, \tau^*, \kappa)$ consists of a set S with $\cdot : S^2 \to S, \tau : S \to S, \tau^* : S \to S, \kappa : 2^S \setminus \{\emptyset\} \to S$ such that (with infix notation for $\cdot$ and superscript notation for τ, τ^*, κ)

A-1 $(S, \cdot)$ is a semigroup.

A-2 $(a \cdot b)^{\tau} = a \cdot (b \cdot a)^{\tau}$ and $(a^n)^{\tau} = a^{\tau}$ for $a, b \in S$ and $n > 0$.

A-3 $(b \cdot a)^{\tau^*} = (a \cdot b)^{\tau^*} \cdot a$ and $(a^n)^{\tau^*} = a^{\tau^*}$ for $a, b \in S$ and $n > 0$.

A-4 For every non-empty subset P of S, every element c in P, every subset Q of P, and every non-empty subset R of $\{P^{\kappa}, a{\cdot}P^{\kappa}, P^{\kappa}{\cdot}b, a{\cdot}P^{\kappa}{\cdot}b \mid a, b \in P\}$, we have $P^{\kappa} = P^{\kappa} \cdot P^{\kappa} = P^{\kappa} \cdot c \cdot P^{\kappa} = (P^{\kappa})^{\tau} = (P^{\kappa} \cdot c)^{\tau} = (P^{\kappa})^{\tau^*} = (c \cdot P^{\kappa})^{\tau^*} = (Q \cup R)^{\kappa}$.

A $\circledast$-*algebra* is a $\oplus$-algebra with a special element 1 where $(S, \cdot, 1)$ is a monoid, $1^{\tau} = 1^{\tau^*} = \{1\}^{\kappa} = 1$ and for all non-empty subsets $P \subseteq S$, $P^{\kappa} = (P \cup \{1\})^{\kappa}$.

A $\oplus$-semigroup naturally induces a $\oplus$-algebra. We simply set $a \cdot b = \pi(ab)$, $a^\tau = \pi(a^\omega), a^{\tau^*} = \pi(a^{\omega^*})$ and $P^\kappa = \pi(P^\eta)$. Similarly a $\circledast$-monoid naturally induces a $\circledast$-algebra with the special element being the neutral element.

Example 3. The $\circledast$-algebra induced by U_1 (recall Example 2) is given below. It plays a crucial role in this work and will also be denoted by U_1.

$\cdot$	1	0	τ	τ^*
1	1	0	1	1
0	0	0	0	0

$$S^\kappa = \begin{cases} 1 & \text{if } S = \{1\} \\ 0 & \text{otherwise} \end{cases}$$

The following is one of the fundamental results of [7, Lemma 3.4 and Theorem 3.11], enabling us to work with $\oplus$-semigroup and $\oplus$-algebra interchangeably as we see fit.

Theorem 1 ([7])**.** *A $\oplus$-semigroup (S, π) induces a unique $\oplus$-algebra. Also, any finite $\oplus$-algebra is induced by a unique $\oplus$-semigroup.*

The proof of Theorem 1 is accomplished in [7] via the novel concept of evaluation trees. Given a $\oplus$-semigroup $(S, \cdot, \tau, \tau^*, \kappa)$, it helps in construction of a unique generalized associativity satisfying map $\pi\colon S^\oplus \to S$ such that (S, π) induces the $\oplus$-algebra $(S, \cdot, \tau, \tau^*, \kappa)$.

Definition 1. An *evaluation tree* over a word $u \in S^\oplus$ is a tree $\mathcal{T} = (T, \iota)$ where T is the set of vertices, and $\iota\colon T \to S$ assigns a value of S to each vertex. Every branch/path of $\mathcal{T}$ is of finite length and every vertex in T is a factor of u. In particular, the root is u. The children of a vertex represent a factorization of the (parent) vertex, and thus the (countable linear) ordering of the children is important. The tree has the following properties:

- A leaf is a singleton letter $a \in S$ such that $\iota(a) = a$.
- Internal nodes have either two or ω or ω^* or η many children.
- If w has two children v_1 followed by v_2, then $w = v_1 v_2$ and $\iota(w) = \iota(v_1) \cdot \iota(v_2)$.
- If w has ω sequence of children $\langle v_1, v_2, \dots \rangle$, then there is an idempotent e such that $e = \iota(v_i)$ for all $i \geq 1$, and $w = \prod_{i \in \omega} v_i$ and $\iota(w) = e^\tau$.

- If w has ω^* sequence of children $\langle \dots, v_{-2}, v_{-1} \rangle$, then there is an idempotent f such that $f = \iota(v_i)$ for all $i \le -1$, and $w = \prod_{i \in \omega^*} v_i$ and $\iota(w) = f^{\tau^*}$.
- If w has children $\langle v_i \rangle_{i \in \eta}$ over η, then $w = \prod_{i \in \eta} v_i$ such that $\prod_{i \in \eta} \iota(v_i)$ is the perfect shuffle of some $E \subseteq S$, and $\iota(w) = E^\kappa$.

The *value* of $\mathcal{T}$ is defined to be $\iota(u)$. Further an ordinal rank can be associated to each node of $\mathcal{T}$ such that the rank of a node is greater than the rank any of its children. This can be used as an induction parameter to reason about any countable word $u \in S^\oplus$. It was shown in [7, Proposition 3.8 and 3.9] that every word u has an evaluation tree and the values of two evaluation trees of u are equal. Setting $\pi(u) = \iota(u)$ creates the necessary map, as it is shown that π defined this way satisfies generalized associativity. Therefore, a $\oplus$-algebra defines the generalized associativity product $\pi \colon S^\oplus \to S$. The correspondence between $\oplus$-semigroups and $\oplus$-algebras permits interchangeability; we implicitly exploit this.

Example 4. Consider the $\circledast$-algebra Gap $= (M, \cdot, \tau, \tau^*, \kappa)$ where $M = \{1, [\,], [\,), (\,], (\,), g\}$, and the operations are defined as follows for M.

$\cdot$	1	[ ]	[ )	( ]	( )	g	τ	τ^*
1	1	[ ]	[ )	( ]	( )	g	1	1
[ ]	[ ]	[ ]	[ )	[ ]	[ )	g	[ )	( ]
[ )	[ )	[ ]	[ )	g	g	g	[ )	( )
( ]	( ]	( ]	( )	( ]	( )	g	( )	( ]
( )	( )	( ]	( )	g	g	g	g	g
g	g	g	g	g	g	g	g	g

$$S^\kappa = \begin{cases} 1 & \text{if } S = \{1\} \\ g & \text{otherwise} \end{cases}$$

It can be easily verified that Gap satisfies the axioms of $\circledast$-algebra. Following our discussion, any countable word $u \in M^\oplus$ is assigned a unique value by this algebra via some evaluation tree for u. For instance consider the evaluation tree for the word $[\,]^\omega [\,]^{\omega^*}$ consisting of a root with two children where the left (resp. right) child represents the word $[\,]^\omega$ (resp. $[\,]^{\omega^*}$); the left (resp. right) child has ω (resp. ω^*) many children $[\,]$ and has value $[\,]^\tau$ (resp. $[\,]^{\tau^*}$). As a result, the value at the root is $[\,]^\tau \cdot [\,]^{\tau^*} = [\,) \cdot (\,] = g$. From our discussion so far, it should be clear that Gap evaluates a word over $\{[\,]\}^\oplus$ to g if and only if the word's underlying linear ordering contains a gap (an ordering α has a gap if it is of the form $\beta_1 + \beta_2$ where β_1 has no maximum element and β_2 has no minimum element).

Now we briefly discuss some natural algebraic notions. Let (S, π) and (S', π') be $\oplus$-semigroups. A morphism from (S, π) to (S', π') is a map $h : S \to S'$ such that, for every $w \in S^{\oplus}$, $h(\pi(w)) = \pi'(\bar{h}(w))$ where $\bar{h}$ is the pointwise extension of h to words. By a slight abuse of notation, we write $h(w)$ for $w \in S^{\oplus}$ to denote $h(\pi(w)) \in S'$. A $\oplus$-language $L \subseteq \Sigma^{\oplus}$ is *recognizable* if there exists a morphism $h\colon (\Sigma^{\oplus}, \prod) \to (S, \pi)$ to a finite $\oplus$-semigroup such that $L = h^{-1}(h(L))$. A $\circledast$-language $L \subseteq \Sigma^{\circledast}$ is *recognizable* if there exists a morphism $h\colon (\Sigma^{\circledast}, \prod) \to (S, \pi)$ to a finite $\circledast$-monoid such that $L = h^{-1}(h(L))$. We'll simply talk about *language* of countable words and let the context explain whether the empty word is being considered or not. Note that these morphisms are completely determined by their restriction to the set Σ, as any map from Σ into S extends to a unique morphism from $\Sigma^{\oplus}$ to (S, π). By the equivalence of finite $\oplus$-semigroup and $\oplus$-algebra, a map from Σ into a $\oplus$-algebra extends to a 'morphism' from $\Sigma^{\oplus}$ into the $\oplus$-algebra, and languages can be naturally recognized via such morphisms.

Example 5. Let $A \subseteq \Sigma$ be a non-empty subset of the alphabet, and L be the set of words that contain an occurence of some letter from A. It is easy to see that the map $h\colon \Sigma \to U_1$ sending $h(a) = 0$ for all $a \in A$, and $h(b) = 1$ for all $b \notin A$ recognizes L as $L = h^{-1}(0)$.

Example 6. Consider the map $h\colon \Sigma \to \text{Gap}$ defined by $h(a) = [\,]$ for all $a \in \Sigma$. The resulting morphism maps any word u to $h(u) = g$ if and only if the domain of the word admits a gap. Consider a word $v = a^{\omega} a^{\omega^*}$ for $a \in \Sigma$. Its pointwise extension under the map h is $\bar{h}(v) = [\,]^{\omega} [\,]^{\omega^*}$. If (Gap, π) is the $\circledast$-monoid that induces the $\circledast$-algebra Gap, then since h extends to a morphism, we have $h(v) = \pi(\bar{h}(v)) = g$ as per the evaluation tree discussion in Example 4.

Remark 1. Let $h\colon \Sigma^{\oplus} \to M$ be a map/morphism into a $\oplus$-algebra. For any word $v \in \Sigma^{\oplus}$, we know its pointwise extension $\bar{h}(v) \in M^{\oplus}$ has an evaluation tree (T, ι). Note that every node in T represents a factor of $\bar{h}(v)$; this factor naturally corresponds to a factor v' of v, that is, the node in T represents $\bar{h}(v')$. Furthermore $h(v')$ is exactly $\iota(\bar{h}(v'))$, the value that ι maps the node to. Therefore the evaluation tree can equivalently be considered over the word $v \in \Sigma^{\oplus}$ with h mapping the word at each node to its evaluation.

Note that (see [10]) any recognizable language L is associated a finite (canonical/minimal) syntactic $\oplus$-semigroup $\mathsf{Syn}(L)$ that divides[1] every $\oplus$-semigroup recognizing L. Further $\mathsf{Syn}(L)$ can be effectively computed from a finite description of L.

We close this section by mentioning the main result of [7].

Theorem 2 ([7])**.** *A language of countable words is recognizable iff it is MSO-definable.*

In the rest of this article we often refer to recognizable languages of countable words as *regular languages* of countable words or simply regular languages.

3. Small Fragments of FO

Our aim is to find algebraic characterizations of interesting logic classes interpreted over countable words. In this section, we focus on two particularly small fragments of first-order logic. First-order logic (FO) over a finite alphabet Σ is a classical logic which can be inductively built using the following operations.

$$\varphi := a(x) \mid x < y \mid \varphi \vee \varphi \mid \neg\varphi \mid \exists x\ \varphi$$

Here $a \in \Sigma$ and φ is any FO formula. We use the letters ϕ, ψ, φ (with or without subscripts) to denote FO formulas, and the letters x, y, z (with or without subscripts) to denote FO variables which represent positions in countable words. We skip the standard semantics.

A sentence is a formula with no free variable. The language of a sentence φ, denoted by $\mathrm{L}(\varphi)$, is the set of all words $u \in \Sigma^\oplus$ that satisfy φ. Let us look at some examples of countable languages definable in FO.

Example 7. The following FO sentence describes the language of all words whose underlying linear ordering is dense and has at least two distinct positions.

$$\exists x \exists y\ x < y \wedge \forall x \forall y\ \big(x < y\big) \Rightarrow \big(\exists z\ x < z < y\big)$$

Example 8. The language of all words containing a gap where the set of letters approaching the gap (arbitrarily closely) from the left is disjoint from

[1] M divides N if M is a homomorphic image of a sub-$\circledast$-semigroup of N

the corresponding set of letters from the right, is definable in FO. In particular, consider the set $\{w_1 w_2 \mid w_1 \in \Sigma^{\circledast}\{a\}^{\oplus}$ has no maximum, and $w_2 \in \{b\}^{\oplus}\Sigma^{\circledast}$ has no minimum$\}$. It is definable in FO by guessing two points x and y in w_1 and w_2 respectively, and expressing the following properties for positions in this interval - (1) all positions are labelled a or b, (2) b labelled positions come after all the a labelled positions, (3) the a-labelled positions do not have a maximum, and (4) the b-labelled positions do not have a minimum.

1. $\varphi_1(x, y) ::= \forall z\ x \leq z \leq y \Rightarrow a(z) \vee b(z)$.
2. $\varphi_2(x, y) ::= \forall z\ (x \leq z \leq y \wedge b(z)) \Rightarrow \neg(\exists z'\ z < z' \leq y \wedge a(z'))$,
3. $\varphi_3(x, y) ::= \forall z\ (x \leq z \leq y \wedge a(z)) \Rightarrow \exists z'\ z < z' < y \wedge a(z')$
4. $\varphi_4(x, y) ::= \forall z\ (x \leq z \leq y \wedge b(z)) \Rightarrow \exists z'\ x < z' < z \wedge b(z')$.

The sentence $\exists x \exists y\ a(x) \wedge b(y) \wedge x < y \wedge \varphi_1(x, y) \wedge \varphi_2(x, y) \wedge \varphi_3(x, y) \wedge \varphi_4(x, y)$ defines the language.

The classical Schützenberger-McNaughton-Papert theorem characterizes FO-definabilty of a regular language of finite words in terms of aperiodicity of its finite syntactic monoid. The survey [20] presents similar decidable characterizations of several interesting small fragments of FO-logic such as FO^1, FO^2, $B(\exists^*)$ – boolean closure of the existential first-order logic. Here we start by identifying algebraic characterizations, over countable words, for FO^1 and $B(\exists^*)$.

3.1. FO *with single variable*

The fragment FO^1 has access to only one variable. We recall that over finite words a regular language is FO^1-definable iff its syntactic monoid is idempotent, that is $x^2 = x$ for any element x, and commutative, that is $x \cdot y = y \cdot x$ for any elements x, y.

Clearly, FO^1 can recognize all words with a particular letter. With a single variable the logic cannot talk about order of positions. This gives an intuition that the syntactic $\oplus$-semigroup of a language definable in FO^1 is commutative. Neither can FO^1 count the number of occurrences of a letter. In short FO^1 can merely detect the presence or absence of a letter.

We say that a $\oplus$-algebra $(M, \cdot, \tau, \tau^*, \kappa)$ is *shuffle-trivial* if it satisfies the following identity: $x_1 \cdot x_2 \cdot \ldots \cdot x_p = \{x_1, \ldots, x_p\}^{\kappa}$. Note that, every element of a shuffle-trivial $\oplus$-algebra is *shuffle-idempotent* (m is a shuffle idempotent if $m^\kappa = m$). From the axioms of a $\oplus$-algebra it easily follows that, m being a shuffle-idempotent implies $m^\tau = m^{\tau^*} = m \cdot m = m$. Furthermore since $x \cdot y = \{x, y\}^\kappa = \{y, x\}^\kappa = y \cdot x$, a shuffle-trivial $\oplus$-algebra is also commutative.

Theorem 3. *Let $L \subseteq \Sigma^\oplus$ be a regular language. The following are equivalent.*

1. *L is recognized by some finite shuffle-trivial $\oplus$-algebra.*
2. *L is a boolean combination of languages of the form $B^\oplus$ where $B \subseteq \Sigma$.*
3. *L is definable in* FO^1.
4. *L is recognized by direct product of U_1s.*
5. *The syntactic $\oplus$-algebra of L is shuffle-trivial.*

Proof.
($1 \Rightarrow 2$) Let L be recognized by a morphism $h\colon \Sigma^\oplus \to (M, \cdot, \tau, \tau^*, \kappa)$ into a finite shuffle-trivial $\oplus$-algebra. Consider an arbitrary word $u \in \Sigma^\oplus$ and let $\mathrm{alph}(u) \subseteq \Sigma$ be the set of letters in the word u, and let $\gamma(u) = \Pi_{a \in \mathrm{alph}(u)} h(a)$ (note that due to commutativity, $\gamma(u)$ is well-defined). We show that $h(u) = \gamma(u)$. The proof is via the evaluation tree (T, h) of the word u. We show by induction on the rank of the nodes in tree (T, h) that $h(v) = \gamma(v)$ for all nodes v in the tree. Consider a node v of the tree.

1. Case v is a letter: The induction hypothesis clearly holds.
2. Case v is a concatenation of words v_1 and v_2: By induction hypothesis $h(v_1) = \gamma(v_1)$ and $h(v_1) = \gamma(v_1)$. Hence we have $h(v) = h(v_1) \cdot h(v_2) = \gamma(v_1) \cdot \gamma(v_2)$. Since $\mathrm{alph}(v) = \mathrm{alph}(v_1) \cup \mathrm{alph}(v_2)$ and all elements of M are idempotents and commute, it is easy to see that $\gamma(v) = \gamma(v_1) \cdot \gamma(v_2)$. Hence $h(v) = \gamma(v)$, and the induction hypothesis holds.
3. Case v is an ω sequence of words $\langle v_1, v_2, \dots \rangle$ such that there exists an $e \in M$ and $h(v_i) = e$ for all $i \geq 1$. Therefore $h(v) = e^\tau$; since in M, $e = e^\tau$, we have $h(v) = e$. We have to show $\gamma(v) = e$.

   Clearly there is a $k \geq 1$ such that $\mathrm{alph}(v_1 v_2 \dots v_k) = \mathrm{alph}(v)$; therefore, denoting $v' = v_1 v_2 \dots v_k$, we know $\gamma(v') = \gamma(v)$. By induction hypothesis and the finite concatenation case seen earlier, we know

$\gamma(v') = h(v') = \Pi_{1\leq i\leq k}h(v_i) = e$. Therefore $\gamma(v) = e = h(v)$, and the induction hypothesis holds in this case.

4. Case v is an ω^* sequence of words: This is symmetric to the case above.

5. Case $v = \prod_{i\in\eta} v_i$ such that $\prod_{i\in\eta} h(v_i)$ is a perfect shuffle of the set $\{b_1, \ldots, b_k\} \subseteq M$. Hence $h(v) = \{b_1, \ldots, b_k\}^\kappa$. By the shuffle-trivial property, we have $h(v) = b_1 \cdot \cdots \cdot b_k$. We have to prove $\gamma(v) = b_1 \cdot \cdots \cdot b_k$.

   Let $l \geq k$ and $j_1, j_2, \ldots, j_l \in \eta$ be such that we get the following: $\{h(v_{j_1}), h(v_{j_2}), \ldots, h(v_{j_l})\} = \{b_1, \ldots, b_k\}$ and $\cup_{1\leq i\leq l}\text{alph}(v_{j_i}) = \text{alph}(v)$. Denoting $v' = v_{j_1} \ldots v_{j_l}$, we therefore get $\gamma(v') = \gamma(v)$, and that $h(v') = \Pi_{1\leq i\leq l}h(v_{j_i})$. Since elements of M commute and are idempotents, we have $h(v') = b_1 \cdot \cdots \cdot b_k$. By the induction hypothesis and finite concatenation case earlier, we can say $\gamma(v') = h(v')$. Hence $\gamma(v) = b_1 \cdot \cdots \cdot b_k$, and the induction hypothesis holds in this case also.

The induction hypothesis, therefore, holds for any word $u \in A^\oplus$. So L is union of equivalence classes defined by the finite index relation $\{(u, v) \mid \text{alph}(u) = \text{alph}(v)\}$. All these classes are boolean combination of languages of the form $B^\oplus$ for some $B \subseteq \Sigma$, as seen below.

$$\{u \mid \text{alph}(u) = B\} = B^\oplus \setminus \left(\bigcup_{b\in B}(B \setminus \{b\})^\oplus\right)$$

($2 \Rightarrow 3$) Note that $B^\oplus$ is expressed by the FO^1 formula $\forall x \vee_{a\in B} a(x)$. The claim follows from boolean closure of FO^1.

($3 \Rightarrow 4$) Due to the restriction of a single variable, any formula $\varphi(x)$ is a boolean combination of atomic letter predicates. Since a position in a word can have exactly one letter, any non-trivial formula $\varphi(x)$ is a disjunction of letter predicates, e.g. $a(x) \vee b(x)$. A language defined by the sentence $\exists x\ (a(x) \vee b(x))$ is recognized by the $\oplus$-algebra U_1 via $h\colon \Sigma \to U_1$ that maps a, b to $0 \in U_1$ and every other letter to $1 \in U_1$. A language defined by boolean combination of such sentences can be recognized by direct products of U_1.

($4 \Rightarrow 5$) The syntactic $\oplus$-algebra of L divides any $\oplus$-algebra that recognizes L; so it divides a direct product of finitely many U_1. It is easily verified that $\oplus$-algebra U_1 is shuffle-trivial. Since these properties are identities, and identities are preserved under direct product and division (see [21]), we get that the syntactic $\oplus$-algebra of L is shuffle-trivial.

(5 $\Rightarrow$ 1) The syntactic $\oplus$-algebra of L is finite because L is a regular language. Also, it is shuffle-trivial by assumption, and a language is always recognized by its syntactic $\oplus$-algebra. So this direction trivially holds. $\square$

3.2. Boolean closure of existential FO

Let us first recall the characterization of $B(\exists^*)$ - the boolean closure of existential FO over finite words. This is precisely the content of the theorem due to Simon [22]. The usual presentation of Simon's theorem refers to piecewise testable languages which are easily seen to be equivalent to $B(\exists^*)$-definable languages. Simon's theorem states that a regular language of finite words is $B(\exists^*)$-definable iff its syntactic monoid is J-trivial. We recall that a monoid M is J-trivial if for all $m, n \in M$, $MmM = MnM$ implies $m = n$. In short, the Green's equivalence relation J on M is the equality relation. We refer to [23] for a detailed study of Green's relations and their use in the proof of Simon's theorem.

The proof of Simon's theorem uses the congruence $\sim_n$, parametrized by $n \in \mathbb{N}$, on finite words Σ^*: for $u, v \in \Sigma^*$, $u \sim_n v$ if u and v have the same set of subwords of length less than or equal to n. Note that $\sim_n$ has finite index.

We fix $n \in \mathbb{N}$ and work with $\sim_n$ defined on countable words $\Sigma^{\circledast}$: for $u, v \in \Sigma^{\circledast}$, $u \sim_n v$ if u and v have the same set of subwords of length less than or equal to n. It is immediate that $\sim_n$ is an equivalence relation on $\Sigma^{\circledast}$ of finite index. We let $S_n = \Sigma^{\circledast}/\sim_n$ denote the finite set of $\sim_n$-equivalence classes. For a word w, $[w]_n$ denotes the $\sim_n$-equivalence class which contains w.

Lemma 1. *There is a natural well-defined product operation $\pi : S_n^{\circledast} \to S_n$ as follows: $\pi\Big(\prod_{i\in\alpha}[w_i]_n\Big) = \big[\prod_{i\in\alpha} w_i\big]_n$. This operation π satisfies the generalized associativity property. As a result, $\mathbf{S_n} = (S_n, 1 = [\varepsilon]_n, \pi)$ is a $\circledast$-monoid.*

Note that the lemma implies that $h_n : \Sigma^{\circledast} \to \mathbf{S_n}$ mapping w to $[w]_n$ is a morphism of $\circledast$-monoids.

Proof. Let $w = \prod_{i\in\alpha} w_i$ and $w' = \prod_{i\in\alpha} w'_i$ where $w_i \sim_n w'_i$ for all $i \in \alpha$. To show π is well defined, we need to show $w \sim_n w'$. Suppose u is a subword of w of length n. We can factorize u as $u = u_1u_2 \dots u_k$ where u_j (for $1 \le j \le k$) is a subword of w_{i_j}. Since $w_{i_j} \sim_n w'_{i_j}$ and $|u_j| \le n$, we have u_j is a subword of w'_{i_j}, and thus u is a subword of w' as well. Therefore, π is well defined.

Next we show that π satisfies the generalized associativity property. Let $u = \prod_{i\in\alpha} u_i$ where $u_i = \prod_{j\in\alpha_i}[v_j]_n$ and α is any countable linear ordering. We have $\pi(u_i) = [\prod_{j\in\alpha_i} v_j]_n$ and hence

$$\pi(\prod_{i\in\alpha}\pi(u_i)) = \left[\prod_{i\in\alpha}(\prod_{j\in\alpha_i} v_j)\right]_n = \pi(u)$$

This completes the proof. $\square$

It is known [21] that a finite monoid $(M,\cdot)$ is J-trivial if and only if it satisfies the (profinite) identities: $x^! = x\cdot x^!$ and $(x\cdot y)^! = (y\cdot x)^!$. Here $x^!$ denotes the unique idempotent in the semigroup generated by x; guarantee of existence and uniqueness of this generated idempotent is a basic result for finite semigroups. We also use the notation $x^!$ for elements of $\circledast$-algebra and it is the idempotent generated by x using the binary concatenation operation. We say that a $\circledast$-algebra is *shuffle-power-trivial* if it satisfies the (profinite) identity: $(x_1\cdot x_2\cdot\ldots\cdot x_p)^! = \{x_1,\ldots,x_p\}^\kappa$. Note that, every idempotent of such a $\circledast$-algebra is a shuffle-idempotent: $x^! = x$ implies $x^\kappa = x$.

Remark 2. Note that in a shuffle-power-trivial algebra, $(x\cdot y)^! = \{x,y\}^\kappa = \{y,x\}^\kappa = (y\cdot x)^!$. Also,

$$x^! = x^\kappa = (x^\kappa)^\tau = (x^!)^\tau = x^\tau = x\cdot x^\tau = x\cdot x^!$$

Thus, a shuffle-power-trivial $\circledast$-algebra is J-trivial. It is also clear that we have $x^! = x^\tau = x^{\tau^*} = x^\kappa$.

Lemma 2. *The $\circledast$-algebra $\mathbf{S_n}$ is shuffle-power-trivial.*

Proof. Let $x_1, x_2, \ldots, x_p \in S_n$. Suppose x_i is the equivalence class of word u_i over Σ. It is easily seen that any n length subword of $\{u_1, u_2, \ldots, u_p\}^\eta$ is also present in $(u_1u_2\ldots u_n)^n$. Therefore $\{x_1, x_2, \ldots, x_p\}^\kappa = (x_1\cdot x_2\ldots x_p)^n$. Since $\{x_1, x_2, \ldots, x_p\}^\kappa$ is idempotent, we get $\{x_1, x_2, \ldots, x_p\}^\kappa = (x_1\cdot x_2\ldots x_p)^!$. $\square$

Theorem 4. *Let $L \subseteq \Sigma^\circledast$ be a regular language. The following are equivalent.*

1. *L is recognized by a finite shuffle-power-trivial $\circledast$-algebra.*
2. *L is recognized by the quotient morphism $h_n : \Sigma^\circledast \to \mathbf{S_n}$ for some n.*
3. *L is definable in $B(\exists^*)$.*

4. The syntactic $\circledast$-algebra of L is shuffle-power-trivial.

Proof.
(1 $\Rightarrow$ 2) Let L be recognized by $h\colon \Sigma^{\circledast} \to \mathbf{M}$ where $\mathbf{M} = (M, 1, \cdot, \tau, \tau^*, \kappa)$ is a finite shuffle-power-trivial $\circledast$-algebra. Since shuffle-power-triviality is preserved in sub-$\circledast$-algebra, we can assume h to be surjective. Consider the restriction of h to the free monoid Σ^* resulting in the induced monoid morphism. We denote it by $h'\colon \Sigma^* \to (M, 1, \cdot)$. By the identities of the $\circledast$-*algebra* $\mathbf{M}$ and its consequences as pointed out in the Remark 2, this morphism is surjective and the monoid $(M, 1, \cdot)$ is J-trivial.

Using the argument from Simon's theorem (see [23, Theorem 3.13]), there exists $n \in \mathbb{N}$, such that $(M, 1, \cdot)$ is a quotient of $\Sigma^*/\sim_n$ and $u \sim_n v$ implies $h'(u) = h'(v)$. We need to 'lift' this result to general countable words. For this we prove that any countable word w has a finite subword $\hat{w}$ such that $w \sim_n \hat{w}$ and $h(w) = h'(\hat{w})$. Let $\mathcal{T} = (T, h)$ be an evaluation tree over w. We prove by induction that for every node v of the tree, there is a finite subword $\hat{v}$ of v with $v \sim_n \hat{v}$ and $h(v) = h'(\hat{v})$.

1. Case v is a letter: The induction hypothesis clearly holds by taking $\hat{v} = v$.

2. Case v is a concatenation of words v_1 and v_2: By induction hypothesis, we have finite subwords $\hat{v_1}$ and $\hat{v_2}$ of v_1 and v_2 respectively such that $\hat{v_1} \sim_n v_1$, $h(v_1) = h'(\hat{v_1})$ and $\hat{v_2} \sim_n v_2$, $h(v_2) = h'(\hat{v_2})$ Note that $\hat{v_1} \sim_n v_1$ and $\hat{v_2} \sim_n v_2$ implies $\hat{v_1}\hat{v_2} \sim_n v_1v_2$. Further, $\hat{v_1}\hat{v_2}$ is a finite subword of v_1v_2 and $h(v) = h(v_1) \cdot h(v_2) = h'(\hat{v_1}) \cdot h'(\hat{v_2}) = h'(\hat{v_1}\hat{v_2})$. This proves the induction hypothesis holds in this case.

3. Case v is an ω sequence of words $\langle v_1, v_2, \dots \rangle$ such that there exists an idempotent $e \in M$ and $h(v_i) = e$ for all $i \geq 1$ and $h(v) = e^{\tau}$. As observed in Remark 2, $e = e^{\kappa} = (e^{\kappa})^{\tau} = e^{\tau}$; therefore we have $h(v) = e$. Because there are only finitely many words of length less than or equal to n, clearly there is a $k \geq 1$ such that $v_1v_2 \dots v_k \sim_n v$. Let us denote $v_1v_2 \dots v_k$ by v'. Note that since e is an idempotent, $h(v') = e = h(v)$. It is now easy to complete the proof by using induction hypothesis for each v_i for $1 \leq i \leq k$ and using the arguments in the concatenation case above.

4. Case v is an ω^* sequence of words: This is symmetric to the case above.

5. Case $v = \prod_{i \in \eta} v_i$ such that $u = \prod_{i \in \eta} h(v_i) \in M^{\oplus}$ is a perfect shuffle of $\{b_1, \ldots, b_k\} \subseteq M$ and $h(v) = \{b_1, \ldots, b_k\}^{\kappa}$. By the shuffle-power-trivial property, we have $h(v) = (b_1 \cdot \ldots \cdot b_k)^!$.

   We claim that there exists a finite subset $X \subset \eta$ such that, with $v' = \prod_{i \in X} v_i$ and $u' = \prod_{i \in X} h(v_i)$, $v \sim_n v'$ and the finite subword u' of u is a large power of the word $b_1 b_2 \ldots b_k$. This would imply $h(v') = (b_1 \cdot \ldots \cdot b_k)^! = h(v)$. We can now apply induction hypothesis on v_i for each $i \in X$ and proceed as in the concatenation case.

   It remains to show the existence of X. We first choose X large enough so that all subwords of v upto length n are represented in v' and then increase X to ensure that u' is of the desired form. This is possible thanks to the fact that u is perfect shuffle of $\{b_1, \ldots, b_k\}$.

Now for any two countable words u and v, if $u \sim_n v$, then $h(u) = h'(\hat{u}) = h'(\hat{v}) = h(v)$ where the middle equality is from the argument used in the proof of Simon's theorem mentioned before. Invoking Lemma 1, it follows that the given morphism h factors through the morphism $h_n : \Sigma^{\circledast} \to \mathbf{S_n}$ that maps u to $[u]_n$.

(2 ⇒ 1) This follows from Lemma 2.

(2 ⇒ 3) Every equivalence class of $\sim_n$ is clearly definable in $B(\exists^*)$.

(3 ⇒ 2) Let L be recognized by the formula $\alpha ::= \exists x_1, \ldots, x_n \varphi(x_1, \ldots, x_n)$. We show that for an $u \sim_n v$, $u \models \alpha$ if and only if $v \models \alpha$. Consider an assignment s which assigns the variables x_is to a position in the domain of u such that $u, s \models \varphi$. Note that since φ is a quantifier free formula it is a boolean combination of formulas of the form $x_i < x_j$, $x_i = x_j$ and $a(x_i)$. Let $X = \{s(x_i) \mid 1 \leq i \leq n\} \subseteq \mathrm{dom}(u)$ be the set of n points which are assigned to the x_is. Since $u \sim_n v$, there is a set $Y \subseteq \mathrm{dom}(v)$ of n points such that $u|_X = v|_Y$. Consider an assignment $\hat{s}$ to variables x_i to positions in Y such that $s(x_i) < s(x_j)$ iff $\hat{s}(x_i) < \hat{s}(x_j)$. Clearly such an assignment satisfies $v, \hat{s} \models \varphi$ since the ordering between the variables and the letter positions are preserved. Therefore we get that $u \models \alpha$ implies $v \models \alpha$. A symmetric argument shows the other direction.

(4 ⇒ 1) This is a trivial observation.

(1 ⇒ 4) This follows from the fact that identities are preserved under division. □

4. Algebraic Products

So far we have provided algebraic characterizations for small fragments of first order logic. Note that the characterizations are of two kinds — decidable characterization in terms of identities (we have given such characterizations for both FO^1 and $B(\exists^*)$), and decompositional characterization where a combination of simple algebraic structures recognize the exact class of language (we have given such a characterization for FO^1). We now move on to characterizing higher logic classes. In [10], decidable characterizations for many interesting logic classes, e.g. FO, have been discovered. So we focus on providing decompositional characterizations instead. Recall that for FO^1, direct product of U_1s provide an exact characterization. However for more expressive logics, direct product is not suitable for getting simple prime algebraic structures, since direct product can only handle boolean combination of languages recognized by individual structures. In the finite words setting, block product is an algebraic product that has played a significant role in providing interesting decompositional characterizations of several logic classes like FO and MSO [15]). Motivated by this, we introduce the block product operation for $\oplus$-semigroups and $\oplus$-algebras, and investigate decompositional characterizations of FO, its subclass FO^2, and also beyond first order logic.

In this section, our aim is to develop a suitable block product operation that is conceptually the right counterpart to the classical notion over monoids and finite words. To achieve this aim, we define the notion of compatible left and right actions on $\oplus$-semigroups and generalize the concept of semidirect product from semigroup theory to this setting. Block product, being a special case of semidirect product, gets defined as a result. A similar development for the block product operation in the classical setting is present in [15]. Finally we establish a result called block product principle which relates language recognized by the block product of two structures in terms of languages recognized by each of the individual structures.

4.1. Actions

Let (M, π) and $(N, \hat{\pi})$ be two $\oplus$-semigroups. Note that the set of all $\oplus$-semigroup morphisms from $(N, \hat{\pi})$ to itself forms a monoid —the endomorphism monoid of N— under function composition. A *left action* of (M, π) on $(N, \hat{\pi})$ is a morphism from M into the endomorphism monoid of N. In other words, it is a map $M \times N \to N$ satisfying the following properties (we denote by mn the element to which the pair (m, n) maps).

B-1 $\pi(m_1 m_2)n = m_1(m_2 n)$

B-2 $m\hat{\pi}(\prod_{i\in\alpha} n_i) = \hat{\pi}(\prod_{i\in\alpha} mn_i)$

If M and N are both $\circledast$-monoids with neutral elements 1 and $\hat{1}$ respectively, then the action is called *monoidal* if, for all $m \in M$, $n \in N$ the following two conditions hold.

C-1 $1n = n$

C-2 $m\hat{1} = \hat{1}$

A right action of M on N is defined symmetrically. M is said to have *compatible* left and right actions on N if the actions commute, or in other words if, for $m, m' \in M$ and $n \in N$, the property $(mn)m' = m(nm')$ is satisfied. We use the notation $m(\prod_{i\in\alpha} n_i)m'$ to denote the natural pointwise extension $\prod_{i\in\alpha} mn_i m'$.

Actions are naturally defined for $\oplus$-algebra as well. Let $(M, \cdot, \tau, \tau^*, \kappa)$ and $(N, +, \hat{\tau}, \hat{\tau}^*, \hat{\kappa})$ be $\oplus$-algebras induced by $\oplus$-semigroups (M, π) and $(N, \hat{\pi})$ respectively. The action requirements can be equivalently stated in terms of algebra operators, e.g. the left action requirements are as follows:

D-1 $(m_1 \cdot m_2)n = m_1(m_2 n)$

D-2 $m(n_1 + n_2) = mn_1 + mn_2$

D-3 $mn^{\hat{\tau}} = (mn)^{\hat{\tau}}$

D-4 $mn^{\hat{\tau}^*} = (mn)^{\hat{\tau}^*}$

D-5 $m\{n_1, \ldots, n_j\}^{\hat{\kappa}} = \{mn_1, \ldots, mn_j\}^{\hat{\kappa}}$

4.2. Semidirect product

We define a bilateral semidirect product of $\oplus$-semigroups (M, π) and $(N, \hat{\pi})$ where M has compatible left and right actions on N. Here onwards we refer to bilateral semidirect product simply as semidirect product. Similarly we refer to compatible left and right actions simply as actions.

Definition 2. Given (M, π) with actions on $(N, \hat{\pi})$, the map $\theta \colon (M \times N)^{\oplus} \to M^{\oplus} \times N^{\oplus}$ associates with any word $u \in (M \times N)^{\oplus}$ two words $v \in M^{\oplus}$ and $w \in N^{\oplus}$ as follows. If $u = \prod_{i\in\alpha}(m_i, n_i)$, then $v = \prod_{i\in\alpha} m_i$ and $w = \prod_{i\in\alpha} \pi(\prod_{j<i} m_j) n_i \pi(\prod_{j>i} m_j)$. See Figure 1.

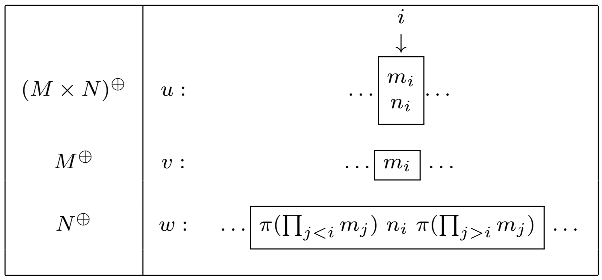

Figure 1: $\theta(u) = (v, w)$

The following lemma states a useful property of the map θ.

Lemma 3. *Consider (M, π) with actions on $(N, \hat{\pi})$. Suppose $u = \prod_{i\in\alpha} u_i \in (M \times N)^{\oplus}$ with $\theta(u) = (v, w)$ and for $i \in \alpha$, $\theta(u_i) = (v_i, w_i)$. Then $v = \prod_{i\in\alpha} v_i$ and $w = \prod_{i\in\alpha} w'_i$ where*

$$w'_i = \pi(\prod_{j<i} v_j) w_i \pi(\prod_{j>i} v_j)$$

Proof. Consider an arbitrary position $l \in \text{dom}(u)$ and let $u[l] = (m, n)$. There exists $i \in \alpha$ such that $l \in \text{dom}(u_i)$. From Definition 2, $v[l] = m = v_i[l]$. In contrast, $w[l] = \pi(v_{<l}) n \pi(v_{>l})$ and $w_i[l] = \pi((v_i)_{<l}) n \pi((v_i)_{>l})$. Note that $v_{<l} = (\prod_{j<i} v_j)(v_i)_{<l}$, and similarly for the suffix $v_{>l}$. Therefore $w[l] = \pi(\prod_{j<i} v_j) w_i[l] \pi(\prod_{j>i} v_j)$ by using generalized associativity of π and action axioms (the axiom B-1 is used for the left action). The lemma follows. $\square$

Definition 3 (Semidirect Product)**.** Given (M, π) with actions on $(N, \hat{\pi})$, their semidirect product $M \ltimes N$ is the pair $(M \times N, \tilde{\pi})$ where $\tilde{\pi} \colon (M \times N)^{\oplus} \to M \times N$ is defined by: for u with $\theta(u) = (v, w)$, we let $\tilde{\pi}(u) = (\pi(v), \hat{\pi}(w))$.

The proof of the following lemma verifies that $M \ltimes N$ is a $\oplus$-semigroup by showing that $\tilde{\pi}$ satisfies the general associativity property.

Lemma 4. *The structure $M \ltimes N = (M \times N, \tilde{\pi})$ is a $\oplus$-semigroup.*

Proof. Let $u = \prod_{i\in\alpha} u_i$ where $u, u_i \in (M \times N)^{\oplus}$. We have to prove $\tilde{\pi}(u) = \tilde{\pi}(\prod_{i\in\alpha} \tilde{\pi}(u_i))$. Rewriting $\prod_{i\in\alpha} \tilde{\pi}(u_i)$ as z, we have to prove $\tilde{\pi}(u) = \tilde{\pi}(z)$.

Suppose $\theta(u) = (v, w)$ and for $i \in \alpha$, $\theta(u_i) = (v_i, w_i)$. Then by Lemma 3, $v = \prod_{i\in\alpha} v_i$ and $w = \prod_{i\in\alpha} w'_i$ where w'_i is as given in the lemma statement. By Definition 3, $\tilde{\pi}(u) = (\pi(v), \hat{\pi}(w))$. Using the generalized associativity properties of π and $\hat{\pi}$, we get $\tilde{\pi}(u) = (\pi(\prod_{i\in\alpha} \pi(v_i)), \hat{\pi}(\prod_{i\in\alpha} \hat{\pi}(w'_i)))$.

Next we analyze the word z. Note that $\mathsf{dom}(z) = \alpha$ and $z[i] = \tilde{\pi}(u_i)$. Further, recall that $\theta(u_i) = (v_i, w_i)$. From Definition 3, we get that $\tilde{\pi}(u_i) = (\pi(v_i), \hat{\pi}(w_i))$. So $z[i] = (\pi(v_i), \hat{\pi}(w_i))$. We now compute $\theta(z)$ using Definition 2. Let $\theta(z) = (z', z'')$. It is easy to see that $z'[i] = \pi(v_i)$. Using this, we see that

$$\begin{aligned} z''[i] &= \pi(\prod_{j<i} \pi(v_j))\hat{\pi}(w_i)\pi(\prod_{j>i} \pi(v_j)) \\ &= \hat{\pi}(\pi(\prod_{j<i} v_j)w_i\pi(\prod_{j>i} v_j)) \\ &= \hat{\pi}(w'_i) \end{aligned}$$

Now we proceed with the computation of $\tilde{\pi}(z)$ by using Definition 3.

$$\begin{aligned} \tilde{\pi}(z) &= (\pi(z'), \hat{\pi}(z'')) \\ &= (\pi(\prod_{i\in\alpha} \pi(v_i)), \hat{\pi}(\prod_{i\in\alpha} \hat{\pi}(w'_i))) \end{aligned}$$

Comparing this with the expression for $\tilde{\pi}(u)$ derived earlier, we see that $\tilde{\pi}(u) = \tilde{\pi}(z)$. This completes the proof. □

Lemma 5. *If M and N are both $\circledast$-monoids and the underlying actions are monoidal, then $M \ltimes N$ is a $\circledast$-monoid.*

Proof. Let M and N have neutral elements 1 and $\hat{1}$ respectively. We prove that $(1, \hat{1})$ is the neutral element of $M \ltimes N$. Consider $u \in (M \times N)^{\circledast}$. Let $\theta(u) = (v, w)$ and $\theta(u_{\neq(1,\hat{1})}) = (v', w')$. If $u[x] = (1, \hat{1})$, then by Definition 2 and by the property of monoidal actions $v[x] = 1$ and $w[x] = \hat{1}$. If $u[x] \neq (1, \hat{1})$, then $v[x] = v'[x]$ and $w[x] = w'[x]$. So $\pi(v) = \pi(v')$ and $\hat{\pi}(w) = \hat{\pi}(w')$. Hence $\tilde{\pi}(u) = \tilde{\pi}(u_{\neq(1,\hat{1})})$. □

Henceforth we work with the assumption that M and N are finite, and turn to the problem of effective construction of semidirect product of finite $\oplus$-algebras. Thanks to Theorem 1, we can restrict our attention to induced $\oplus$-algebras. Towards this, let $(M, \cdot, \tau, \tau^*, \kappa)$ and $(N, +, \hat{\tau}, \hat{\tau}^*, \hat{\kappa})$ be $\oplus$-algebras induced by $\oplus$-semigroups (M, π) and $(N, \hat{\pi})$ respectively. Further, let $M \ltimes N = (M \times N, \tilde{\cdot}, \tilde{\tau}, \tilde{\tau}^*, \tilde{\kappa})$ denote the $\oplus$-algebra induced by $M \ltimes N = (M \times N, \tilde{\pi})$.

The following lemma says that the binary operator $\tilde{\cdot}$ of $M \ltimes N$ can be expressed using the binary operators $\cdot$ (of M) and $+$ (of N). It follows easily from the definition of the *induced* operator $\tilde{\cdot}$ from $\tilde{\pi}$. We skip the proof as this is same as the classical case.

Lemma 6. *The operator* $\tilde{\cdot}$ *can be defined as follows:* $(m_1, n_1) \mathbin{\tilde{\cdot}} (m_2, n_2) = (m_1 \cdot m_2, n_1 m_2 + m_1 n_2)$.

An easy consequence of the previous lemma is that if (m, n) is an idempotent element of $M \ltimes N$ then m is also an idempotent element of M.

Now we focus on the unary operators $\tilde{\tau}$ and $\tilde{\tau}^*$. In view of the second axiom in the definition of a $\oplus$-algebra, it suffices to show that these operators can be computed at idempotent elements of $M \ltimes N$ in terms of the algebra operators of M and N.

Lemma 7. *Let* (e, n) *be an idempotent element of* $M \ltimes N$*. Then* $(e, n)^{\tilde{\tau}} = (e^{\tau}, ne^{\tau} + (ene^{\tau})^{\hat{\tau}})$*, and* $(e, n)^{\tilde{\tau}^*} = (e^{\tau^*}, (e^{\tau^*}ne)^{\hat{\tau}^*} + e^{\tau^*}n)$.

Proof. We present the proof only for $\tilde{\tau}$. By definition of the induced operator $\tilde{\tau}$, $(e, n)^{\tilde{\tau}} = \tilde{\pi}(u)$ where $u = (e, n)^{\omega}$ is the ω-word over the domain $(\mathbb{N}, <)$ such that every position is mapped to (e, n). We first compute $\theta(u) = (v, w)$ according to the Definition 2. It is easy to see that $v = e^{\omega}$ and w is the ω-word whose first position is mapped to ne^{τ} and all other positions are mapped to ene^{τ}. As a result, $\pi(v) = e^{\tau}$ and $\hat{\pi}(w) = ne^{\tau} + (ene^{\tau})^{\hat{\tau}}$. The proof now follows by observing that $\tilde{\pi}(u) = (\pi(v), \hat{\pi}(w))$. □

Finally, the next lemma shows that the operator $\tilde{\kappa}$ of $M \ltimes N$ can be computed using the algebra operators of M and N.

Lemma 8. *The operator* $\tilde{\kappa}$ *can be defined as follows:*

$$\{(m_1, n_1), \ldots, (m_p, n_p)\}^{\tilde{\kappa}} = (m, \{mn_1m, \ldots, mn_pm\}^{\hat{\kappa}})$$

where $m = \{m_1, \ldots, m_p\}^{\kappa}$.

Proof. Let $S = \{(m_1, n_1), \ldots, (m_p, n_p)\}$. Then if u is the perfect shuffle of S, that is, if $u = S^{\eta}$, then $\tilde{\pi}(u) = S^{\kappa}$. Consider $\theta(u) = (v, w)$. We claim v is the perfect shuffle of the set $S_1 = \{m_1, \ldots, m_p\}$. Indeed for any two points $x < y$ in $\mathrm{dom}(v)$, if suppose m_1 is not present, then between the same points in $\mathrm{dom}(u)$ the element (m_1, n_1) is not present. Therefore $v = S_1^{\eta}$, and $\pi(v) = S_1^{\kappa} = m$ (say). Furthermore for any point i in $\mathrm{dom}(v)$, the prefix $v_{<i}$ and the suffix $v_{>i}$ are both perfect shuffles of S_1; so $\pi(v_{<i}) = \pi(v_{>i}) = m$. This implies w is the perfect shuffle of the set $S_2 = \{mn_1m, \ldots, mn_pm\}$. The result follows as $\hat{\pi}(w) = S_2^{\kappa}$, and $\tilde{\pi} = (\pi(v), \hat{\pi}(w)) = (m, S_2^{\kappa})$. □

We now present an example of a semidirect product construction.

Example 9. Consider $M = \mathrm{U}_1$ acting on $N = \mathrm{U}_1$ with a trivial left action and a non-trivial monoidal right action where $0 \in M$ maps everything in N to $1 \in N$. The $\circledast$-algebra $\mathcal{S} = \mathrm{U}_1 \ltimes \mathrm{U}_1$ is given in Figure 2. We write the element (i, j) as ij in this example.

$\cdot$	11	10	00	01	τ	τ^*
11	11	10	00	01	11	11
10	10	10	00	01	10	10
00	00	00	00	01	01	00
01	01	00	00	01	01	01

$$S^\kappa = \begin{cases} 11 & \text{if } S = \{11\} \\ 01 & \text{if } S \cap \{00, 01\} \neq \emptyset \\ 10 & \text{otherwise} \end{cases}$$

Figure 2: The $\circledast$-algebra $\mathcal{S} = U_1 \ltimes U_1$

Example 10. Let $\Sigma = \{a, b\}$. Consider the language L of all words which contains the letter b, and has a non-empty suffix purely consisting of a's, that is, $L = \Sigma^\circledast \cdot \{b\} \cdot \Sigma^\circledast \cdot \{a\}^\oplus$. The morphism $h \colon \Sigma^\oplus \to \mathcal{S}$ such that $h(a) = 10$ and $h(b) = 01$ recognizes L as $L = h^{-1}(00)$.

4.3. Block Product

Let (M, π) and $(N, \hat{\pi})$ be two $\oplus$-semigroups. Recall that M^1 is the $\circledast$-monoid associated to M. The set $N^{M^1 \times M^1}$ of all functions from $M^1 \times M^1$ into N also forms a $\oplus$-semigroup under the componentwise product. This $\oplus$-semigroup can be simply viewed as the direct product of $|M^1| \times |M^1|$ copies of N. Reusing the operation $\hat{\pi}$ of $(N, \hat{\pi})$, we denote this $\oplus$-semigroup by $(K, \hat{\pi})$ with underlying set $K = N^{M^1 \times M^1}$

The block product of M and N is denoted by $M \Box N$ and is the semidirect product $M \ltimes K$ (with underlying set $M \times K$) with respect to the *canonical* 'actions' (the following lemma proves that these are indeed compatible left and right actions): for $m \in M$ and $f \in K$,

$$(mf)(m_1, m_2) = f(m_1 m, m_2)$$

$$(fm)(m_1, m_2) = f(m_1, m m_2)$$

Lemma 9. *Given $\oplus$-semigroups (M, π) and $(N, \hat{\pi})$, consider the maps $M \times N^{M^1 \times M^1} \to N^{M^1 \times M^1}$ defined by $(mf)(m_1, m_2) = f(m_1 m, m_2)$ and $N^{M^1 \times M^1} \times$*

$M \to N^{M^1 \times M^1}$ defined by $(fm)(m_1, m_2) = f(m_1, mm_2)$. These are compatible left and right actions of (M, π) on $(N^{M^1 \times M^1}, \hat{\pi})$. They are also monoidal if M and N are both $\circledast$-monoids.

Proof. We focus only on the left action. Note that

$$\begin{aligned}(m'(mf))(m_1, m_2) &= (mf)(m_1 m', m_2) \\ &= f(m_1 m' m, m_2) \\ &= ((m'm)f)(m_1, m_2)\end{aligned}$$

Hence $m'(mf) = (m'm)f$, thus proving the first axiom. For the second axiom, note

$$\begin{aligned}(m(\prod_{i \in \alpha} f_i))(m_1, m_2) &= (\prod_{i \in \alpha} f_i)(m_1 m, m_2) \\ &= \prod_{i \in \alpha}(f_i(m_1 m, m_2)) \\ &= \prod_{i \in \alpha}(mf_i(m_1, m_2)) \\ &= (\prod_{i \in \alpha} mf_i)(m_1, m_2)\end{aligned}$$

So $m(\prod_{i \in \alpha} f_i) = \prod_{i \in \alpha} mf_i$, thus proving the second axiom. If M has neutral element 1, then $(1f)(m_1, m_2) = f(m_1, m_2)$ which means $1f = f$. If N has neutral element $1'$, then the neutral element g of K is the constant function to $1'$. Clearly, $mg = g$. Thus the left action is monoidal if (M, π) and $(N, \hat{\pi})$ are $\circledast$-monoids.

The proof for the right action is symmetrical. We now establish the compatibility of these two actions.

$$((mf)m')(m_1, m_2) = (mf)(m_1, m'm_2) = f(m_1 m, m'm_2)$$

$$(m(fm'))(m_1, m_2) = (fm')(m_1 m, m_2) = f(m_1 m, m'm_2)$$

Therefore $(mf)m' = m(fm')$, that is, the actions commute and are compatible. This completes the proof. □

4.4. Block Product Principle

In this subsection, we state and prove the block product principle. Roughly speaking the block product principle allows to express the formal languages recognized by the block product $M \Box N$ in terms of languages recognized by M and N.

Fix a finite alphabet Σ. As $\Sigma^{\oplus}$ is a free $\oplus$-semigroup, a morphism from $\Sigma^{\oplus}$ to $M \Box N = M \ltimes K$ is simply given (determined) by a map $h : \Sigma \to M \times K$. Sometimes we'll denote its pointwise extension $\bar{h} : \Sigma^{\oplus} \to (M \times K)^{\oplus}$ also by h. Further, composing this with the countable product $\tilde{\pi}$ of $M \ltimes K$ results into a *morphism* which, to a word $u \in \Sigma^{\oplus}$, associates the element $\tilde{\pi}(\bar{h}(u)) \in M \times K$. This morphism may also be denoted by h (that is, $h(u)$ may simply equal $\tilde{\pi}(\bar{h}(u))$). The context will make it clear as to which interpretation of 'h' applies. These slight abuses of notations are used several times in what follows in order to keep the notation simple and improve readability.

Similar to the finite words case, the block product principle over countable words crucially utilises a sequential transducer induced by morphisms from the free $\oplus$-semigroup.

Definition 4. Let $\varphi\colon \Sigma^{\oplus} \to (M, \pi)$ be a morphism. The sequential transducer σ_φ associated with this morphism is a domain-preserving letter-to-letter transducer of type $\sigma_\varphi\colon \Sigma^{\oplus} \to (M^1 \times \Sigma \times M^1)^{\oplus}$ and is defined as follows. For any word $u \in \Sigma^{\oplus}$, and for any $x \in \mathrm{dom}(u)$,

$$\sigma_\varphi(u)[x] = (\varphi(u_{<x}), u[x], \varphi(u_{>x}))$$

As mentioned earlier $\mathrm{dom}(\sigma_\varphi(u)) = \mathrm{dom}(u)$.

Remark 3. If the prefix $u_{<x}$ (resp. suffix $u_{>x}$) is the empty word in Definition 4, then we use the neutral element of M^1 in place of $\varphi(u_{<x})$ (resp. $\varphi(u_{>x})$).

Next, given a morphism from a free $\oplus$-semigroup into a block product $\oplus$-semigroup, we define two naturally arising morphisms into the individual $\oplus$-semigroups of the block product.

Definition 5. Let $h : \Sigma^{\oplus} \to M \Box N$ be a morphism and let $(m_a, f_a) = h(a)$ for each $a \in \Sigma$. We define the map/morphism $h_{\mathbb{1}} : \Sigma \to M$ by letting $h_{\mathbb{1}}(a) = m_a$ for each letter a. We also define the map/morphism $h_{\mathbb{2}} : (M^1 \times \Sigma \times M^1) \to N$ as: for $(m_1, a, m_2) \in (M^1 \times \Sigma \times M^1)$, we have $h_{\mathbb{2}}((m_1, a, m_2)) = f_a(m_1, m_2)$.

Going ahead, given a word $u' \in (M^1 \times \Sigma \times M^1)^\oplus$ and $m_1, m_2 \in M$, we define $m_1 u' m_2$ to be the word (with the same domain as u') such that for a position x with $u'[x] = (m'_1, a, m'_2)$, $(m_1 u' m_2)[x] = (m_1 m'_1, a, m'_2 m_2)$.

Now we are ready to state a key technical lemma which will help us establish the block product principle.

Lemma 10. *Consider a morphism $h\colon \Sigma^\oplus \to M \Box N = M \ltimes K$. For $u \in \Sigma^\oplus$, we have $h(u) = (m, f)$ if and only if $h_{\mathbb{1}}(u) = m$ and for all $m_1, m_2 \in M^1$, we have $h_{\mathbb{2}}(m_1 \sigma(u) m_2) = f(m_1, m_2)$ where σ is the sequential transducer associated to $h_{\mathbb{1}}$.*

Proof. Fix $u \in \Sigma^\oplus$ and $u' = \sigma(u)$. Let $h(u) \in (M \times K)^\oplus$ be the image of the pointwise extension of h applied to u. The words $h_{\mathbb{1}}(u) \in M^\oplus$ and $h_{\mathbb{2}}(u') \in N^\oplus$ are defined similarly. Observe that, for a position x of u, with $u[x] = a$ and $h(a) = (m_a, f_a)$, $h(u)[x] = (m_a, f_a)$, $h_{\mathbb{1}}(u)[x] = m_a$, $u'[x] = (h_{\mathbb{1}}(u_{<x}), a, h_{\mathbb{1}}(u_{>x}))$ and $h_{\mathbb{2}}(u')[x] = f_a(h_{\mathbb{1}}(u_{<x}), h_{\mathbb{1}}(u_{>x}))$. See Figure 3.

		x ↓	⇝ (evaluation)
	u :	$\dots \boxed{a} \dots$	
$h : \Sigma \to M \Box N$	$h(u)$:	$\dots \boxed{(m_a, f_a)} \dots$	(m, f)
$h_{\mathbb{1}} : \Sigma \to M$	$h_{\mathbb{1}}(u)$:	$\dots \boxed{m_a} \dots$	m
$\sigma : \Sigma^\oplus \to (M^1 \times \Sigma \times M^1)^\oplus$	$u' = \sigma(u)$:	$\dots \boxed{h_{\mathbb{1}}(u_{<x}), a, h_{\mathbb{1}}(u_{>x})} \dots$	
$h_{\mathbb{2}} : (M^1 \times \Sigma \times M^1) \to N$	$h_{\mathbb{2}}(u')$:	$\dots \boxed{f_a(h_{\mathbb{1}}(u_{<x}), h_{\mathbb{1}}(u_{>x}))} \dots$	$f(1,1)$

Figure 3: The block product operational view

Consider the map $\theta : (M \times K)^\oplus \to M^\oplus \times K^\oplus$ from Lemma 3 (with K playing the role of N in the statement). Let $\theta(h(u)) = (v, w)$. Observe that $v \in M^\oplus$ and $w \in K^\oplus$. It is straightforward to check that $v = h_{\mathbb{1}}(u)$. Further, by the definition of θ, for a position x of u, with $h(u)[x] = (m_a, f_a)$, $w[x] = h_{\mathbb{1}}(u_{<x}) f_a h_{\mathbb{1}}(u_{>x})$.

Now we relate the word $w \in K^\oplus$ with $\sigma(u) \in (M^1 \times \Sigma \times M^1)^\oplus$. Towards this, consider the projection morphisms: for $m_1, m_2 \in M^1$, $\Pi_{m_1,m_2} : K \to N$ defined as $\Pi_{m_1,m_2}(g) = g(m_1, m_2)$. As expected, the pointwise extensions of Π_{m_1,m_2} are also denoted by Π_{m_1,m_2}.

For further analysis, fix a choice of $m_1, m_2 \in M^1$. Let x be a position with $u[x] = a$ and $h(a) = (m_a, f_a)$. As observed earlier $w[x] =$

$h_{\mathbb{1}}(u_{<x})f_a h_{\mathbb{1}}(u_{>x}) \in K$, and $u'[x] = (h_{\mathbb{1}}(u_{<x}), a, h_{\mathbb{1}}(u_{>x})) \in M^1 \times \Sigma \times M^1$. Clearly $m_1 u' m_2[x] = (m_1 h_{\mathbb{1}}(u_{<x}), a, h_{\mathbb{1}}(u_{>x})m_2)$.

We proceed further with some simple calculations.

$$\begin{aligned} \Pi_{m_1,m_2}(w[x]) &= (h_{\mathbb{1}}(u_{<x})f_a h_{\mathbb{1}}(u_{>x}))\,(m_1, m_2) \\ &= f_a(m_1 h_{\mathbb{1}}(u_{<x}), h_{\mathbb{1}}(u_{>x})m_2) \end{aligned}$$

$$\begin{aligned} h_{\mathbb{2}}(m_1 u' m_2[x]) &= h_{\mathbb{2}}\left((m_1 h_{\mathbb{1}}(u_{<x}), a, h_{\mathbb{1}}(u_{>x})m_2)\right) \\ &= f_a(m_1 h_{\mathbb{1}}(u_{<x}), h_{\mathbb{1}}(u_{>x})m_2) \end{aligned}$$

This reveals that for each position x, $\Pi_{m_1,m_2}(w[x]) = h_{\mathbb{2}}(m_1 u' m_2[x])$. Thanks to the fact that both $\Pi_{m_1,m_2}(w)$ and $h_{\mathbb{2}}(m_1 u' m_2)$ are defined pointwise, we have $\Pi_{m_1,m_2}(w) = h_{\mathbb{2}}(m_1 u' m_2)$. We let f denote the evaluation of w in K and exploit the fact that both Π_{m_1,m_2} and $h_{\mathbb{2}}$ are *morphisms* to conclude that, for $m_1, m_2 \in M$, $f(m_1, m_2) = h_{\mathbb{2}}(m_1 u' m_2) \in N$.

With $h_{\mathbb{1}}(u) = m$, the proof of the proposition is now immediate by Definition 3 which asserts that $h(u) = (m, f)$. □

We now use this lemma to derive the following result often referred to as *the block product principle* (see [23, 24] for the related wreath product principle in finite case).

Theorem 5 (Block Product Principle)**.** *Let $L \subseteq \Sigma^{\oplus}$ be recognized by $h : \Sigma^{\oplus} \to M \Box N$ via a subset F. Let $h_{\mathbb{1}} \colon \Sigma^{\oplus} \to M$ be the induced projection morphism, and let $\sigma \colon \Sigma^{\oplus} \to (M^1 \times \Sigma \times M^1)^{\oplus}$ be the sequential letter-to-letter transducer associated to $h_{\mathbb{1}}$. Then L can be expressed as a finite union*

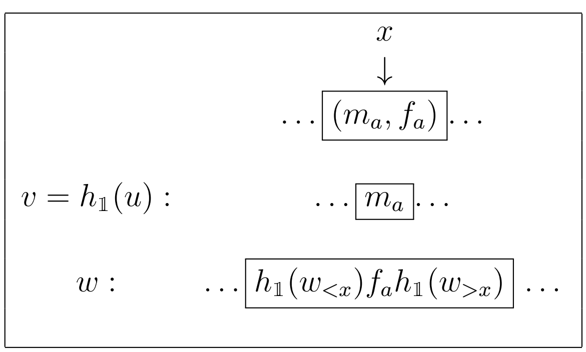

Figure 4: $\theta : (M \times K)^{\circledast} \to M^{\circledast} \times K^{\circledast}$ and $\theta(u) = (v, w)$

of languages of the form $L_1 \cap (\bigcap_{i,j} \sigma^{-1}(L_{ij}))$ where L_1 and L_{ij} are recognized by M and N respectively, for $1 \le i, j \le |M^1|$.

Conversely let $g_1 \colon \Sigma^{\oplus} \to P$ be a morphism, and let $\theta \colon \Sigma^{\oplus} \to (P^1 \times \Sigma \times P^1)^{\oplus}$ be the letter-to-letter transducer associated to it. If $X \subseteq (P^1 \times \Sigma \times P^1)^{\oplus}$ is recognized by some $\oplus$-semigroup Q, then $\theta^{-1}(X)$ is recognized by $P \Box Q$.

Proof. Consider an element $(m, f) \in M \Box N$. By Lemma 10, for $u \in \Sigma^{\oplus}$, $h(u) = (m, f)$ iff $h_{\mathbb{1}}(u) = m$ and $h_{\mathbb{2}}(m_1 \sigma(u) m_2) = f(m_1, m_2)$ for all $m_1, m_2 \in M^1$.

Next, for $1 \le i, j \le |M^1|$, we define the maps/morphisms $h_{ij} : (M^1 \times \Sigma \times M^1) \to N$ as follows: $h_{ij}((m_1, a, m_2)) = h_{\mathbb{2}}((m_i m_1, a, m_2 m_j))$. It is easy to see that, for any word $u' \in (M^1 \times \Sigma \times M^1)^{\oplus}$, $h_{ij}(u') = h_{\mathbb{2}}(m_i u' m_j)$.

As a consequence, we get

$$L = \bigcup_{(m,f) \in F} \left({h_{\mathbb{1}}}^{-1}(m) \cap \left(\bigcap_{i,j} \sigma^{-1}({h_{ij}}^{-1}(f(m_i, m_j))) \right) \right)$$

This completes the proof for one direction.

For the converse, suppose $X \subseteq (P^1 \times \Sigma \times P^1)^{\oplus}$ is recognized by some morphism $g_2 \colon (P^1 \times \Sigma \times P^1)^{\oplus} \to Q$ via subset $F' \subseteq Q$. Consider the map/morphism $g \colon \Sigma^{\oplus} \to P \Box Q$ defined by $g(a) = (g_1(a), \{(m_1, m_2) \mapsto g_2(m_1, a, m_2)\})$. For any word $u \in \Sigma^{\oplus}$, we know $u \in \theta^{-1}(X)$ iff $\theta(u) \in X$ iff $g_2(\theta(u)) \in F'$. It is easy to verify that the map/morphism $g_{\mathbb{2}}$ induced by g (cf. Definition 5) is same as g_2. Therefore, by Lemma 10, $g_2(\theta(u)) = q(1, 1)$ if $g(u) = (p, q)$. As a consequence, we get

$$X = g^{-1}(\{(p, q) \in P \Box Q \mid q(1, 1) \in F'\})$$

This completes the proof. □

Example 11. Let $\Sigma = \{a, b\}$. Recall (see Example 5) that U_1 recognizes the language L_1 of words in which there is at least one occurence of a. We show that $U_1 \Box U_1$ recognizes the language L of words where there is exactly one occurence of a. Let $h \colon \Sigma^{\oplus} \to U_1$ be the morphism recognizing the language L_1 as $L_1 = h^{-1}(0)$, and let $\sigma \colon \Sigma^{\oplus} \to (U_1 \times \Sigma \times U_1)^{\oplus}$ be the canonical transdsucer associated to it. If $\sigma(w)[i] = (1, a, 1)$, then by definition of the transducer, we can say $w[i] = a$, $w_{<i} \notin L_1$ and $w_{>i} \notin L_1$. Consider the language $L_2 \subseteq (U_1 \times \Sigma \times U_1)^{\oplus}$ of words in which there is at least one occurence of the letter $(1, a, 1)$ (note that by the behaviour of σ, there can be at most one such letter in the transducer output). Clearly L_2 is recognized by U_1 and $L = \sigma^{-1}(L_2)$. Therefore by proposition 5, L is recognized by $U_1 \Box U_1$.

5. Block Product Closures and FO² Logic

Having set up the block product operation, we now present a characterization using it. The two variable fragment of first order logic, FO^2, has been studied extensively, particularly in the context of finite words. A block product characterization in terms of U_1s is established in [16] over finite words. In this section, we show that the countable counterpart of the result holds as well. Before stating the characterization, we need to introduce some closures of block product iterations, and their properties.

5.1. Iterated and Weakly Iterated Block Product

Block product of $\oplus$-semigroup is not associative. This is easily evidenced by a cardinality argument, for instance between $(U_1 \Box U_1) \Box U_1$ and $U_1 \Box (U_1 \Box U_1)$. Thus given a list of $\oplus$-semigroups, the order of product (equivalently the nesting of brackets) varies the resulting structure.

We define two particular nestings which will be of interest to us. For a set P of $\oplus$-semigroups, an *iterated block product* is defined inductively as follows:

1. S is an iterated block product for any $S \in P$.
2. If S' is an iterated block product, then $S' \Box S$ is an iterated block product for any $S \in P$.

The set of all iterated block products of a set P is denoted by $\Box^* P$. For a singleton set, we drop the set notation. For instance, $(U_1 \Box U_1) \Box U_1 \in \Box^* U_1$. For a sequence of $\oplus$-semigroups $S_1, \ldots, S_k$, we denote its iterated block product $(\ldots((S_1 \Box S_2) \Box S_3) \ldots) \Box S_k$ by $\Box(S_1, S_2, \ldots, S_k)$.

The following lemma states that direct product of iterated block products is simulated by an iterated block product of the same constituents. The proof follows the corresponding one for classical semigroups (see [15, Appendix A.4]).

Lemma 11. *If $M_1 \prec \Box(S_1, \ldots, S_k)$ and $M_2 \prec \Box(S'_1, \ldots, S'_l)$, then*

$$M_1 \times M_2 \prec \Box(S_1, \ldots, S_k, S'_1, \ldots, S'_l)$$

The other important nesting is *weakly iterated block product.* Given a set P of $\oplus$-semigroups, it is defined inductively as follows:

1. S is a weakly iterated block product for any $S \in P$.
2. If S' is a weakly iterated block product, then $S \Box S'$ is a weakly iterated block product for any $S \in P$.

The set of all weakly iterated block products of a set P is denoted by $\Box_w^* P$. For instance, $U_1 \Box (U_1 \Box U_1) \in \Box_w^* U_1$. For a sequence of $\oplus$-semigroups $S_1, \ldots, S_k$, we denote $S_1 \Box (S_2 \Box \ldots (S_{k-1} \Box S_k) \ldots)$, its weakly iterated block product, by $\Box_w(S_1, S_2, \ldots, S_k)$.

Lemma 12. *For any $\oplus$-semigroups $S_1, \ldots, S_k$, the following holds*

$$(S_1 \times \ldots \times S_{k-1}) \Box S_k \prec \Box_w(S_1, \ldots, S_k)$$

Proof. This follows from a simple inductive argument on k. For $k = 3$, consider the map $h\colon (S_1 \times S_2) \Box S_3 \to S_1 \Box (S_2 \Box S_3)$ defined by: for any $((s_1, s_2), f) \in (S_1 \times S_2) \Box S_3$, its image is (s_1, f') where for any $s, s' \in S_1$, and any $s_2', s_2'' \in S_2$

$$f'(s, s') = (s_2, \{(s_2', s_2'') \mapsto f((s, s_2'), (s', s_2''))\})$$

It can be verified that h is an injective morphism, thus showing $(S_1 \times S_2) \Box S_3$ is isomorphic to a sub-$\circledast$-algebra of $\Box_w(S_1, S_2, S_3)$.

So for $k <= 3$, the statement holds. Assuming it holds for $k - 1$, we get

$$\begin{aligned}(S_1 \times \ldots \times S_{k-1}) \Box S_k &\prec (S_1 \times \ldots S_{k-2}) \Box (S_{k-1} \Box S_k) \\ &\prec \Box_w(S_1, \ldots, S_{k-2}, (S_{k-1} \Box S_k)) \\ &= \Box_w(S_1, \ldots, S_k)\end{aligned}$$

This completes the proof. $\Box$

5.2. FO with two variables

We now consider the two variable fragment FO^2 of first order logic. Over finite words, FO^2 can talk about occurrence of letters and also about the order in which they appear. Over countable linear orderings, it can also say that there is no maximum position. For example, the following formula states that every position is labelled by a and there is no maximum position.

$$\big(\forall x\ \exists y\ x < y\big) \land \big(\forall x\ a(x)\big)$$

Analogously, FO^2 can also talk about words with no minimum position. However, the two variable fragment is not as expressive as full first order. FO^2 satisfies a downward property (similar to Löwenheim-Skolem downward theorem for first order logic): a satisfiable FO^2 formula has a scattered satisfying model [11]. Therefore, the language in Example 7, which says the linear ordering is dense and has at least two distinct positions, is not definable in FO^2. We now present a decompositional characterization of FO^2 languages. The proof follows the one for finite words in [16].

Theorem 6. *A language is definable in* FO^2 *if and only if it is recognised by a weakly iterated block product of* U_1*.*

Proof. The right to left inclusion is via induction on the number of blocks of U_1s. First, observe that languages recognized by a single U_1 can be defined in FO^2. For the induction step, we utilise Theorem 5, the block product principle. Let the hypothesis hold for algebra $M \in \Box_w^* \mathrm{U}_1$. We show that a language L recognized by some morphism $h : \Sigma \to \mathrm{U}_1 \Box M$ can be defined in FO^2. Let $\sigma : \Sigma^{\oplus} \to (\mathrm{U}_1 \times \Sigma \times \mathrm{U}_1)^{\oplus}$ be the transducer associated with the induced morphism $h_{\mathbb{1}} : \Sigma \to U_1$. From the block product principle, L can be expressed as a finite boolean combination of languages of the form L_1 and $\sigma^{-1}(L_2)$ where L_1 and L_2 are recognized by U_1 and M respectively. By the induction hypothesis both L_1 and L_2 are FO^2 definable. So it suffices to show that for an FO^2 language L_2 over the alphabet $(\mathrm{U}_1 \times \Sigma \times \mathrm{U}_1)$ the language $\sigma^{-1}(L_2)$ is also FO^2 definable. This can be shown via structural induction on formula over the decorated alphabet; the base case is the non-trivial case. The following formula accepts $\sigma^{-1}(L_2)$ if L_2 is defined by the formula $(0, a, 1)(x)$.

$$a(x) \wedge \big(\exists y\ y < x \wedge \bigvee_{h_1(b)=0} b(y)\big) \wedge \big(\forall y\ y > x \Rightarrow \bigvee_{h_1(c)=1} c(y)\big)$$

Note that we used only two variables for the above translation. The other base cases are similar. We apply this translation inductively for other formulas.

Now we show the left to right inclusion of the proof. First we note the following observation. Consider $\wp(\Sigma)$, the powerset of the alphabet, as a $\circledast$-monoid where any word $u \in (\wp(\Sigma))^{\oplus}$ is evaluated to the set of letters present in u. Notice that $\wp(\Sigma)$ is essentially the direct product of $|\Sigma|$-many U_1s. There exists a canonical morphism $g : \Sigma^{\oplus} \to \wp(\Sigma)$

such that $g(w) = \{a \mid \text{the letter } a \text{ occurs in } w\}$. The transducer associated with g is $\sigma : \Sigma^{\oplus} \to (\wp(\Sigma) \times \Sigma \times \wp(\Sigma))^{\oplus}$ where, for a word w, we have $\sigma(w)[i] = (g(w_{<i}), w[i], g(w_{>i}))$ for every position i in $\text{dom}(w)$. Observe that the word $\sigma(w)$ carries, at every position i, the information about the set of letters occuring to the left (as well as right) of i in w.

It is shown in [16] that FO^2 has a "normal form" where the quantifier at the maximum depth along with its scope is of the form $\exists x(a(x) \wedge x < y)$ or $\exists x\ (a(x) \wedge x > y)$. Our proof is via induction on the quantifier depth and the number of quantifiers at the maximum depth.

Consider a FO^2 sentence ϕ in its normal form. Consider a subformula $\exists x(a(x) \wedge x < y)$ at its maximum quantifier depth. We convert the formula ϕ into a formula ϕ' over $\wp(\Sigma) \times \Sigma \times \wp(\Sigma)$ as follows. We substitute the chosen subformula $\exists x(a(x) \wedge x < y)$ by a disjunction of letter formulas $(\Sigma_1, b, \Sigma_2)(y)$ where $\Sigma_1, \Sigma_2 \subseteq \Sigma$, $b \in \Sigma$, and $a \in \Sigma_1$. All remaining instances of letter formula $c(x)$ is substituted by disjunction of letter formulas $(\Sigma'_1, c, \Sigma'_2)(x)$ where $\Sigma'_1, \Sigma'_2 \subseteq \Sigma$. It is easy to verify by structural induction on FO^2 formulas that $w \models \phi$ if and only if $\sigma(w) \models \phi'$. In ϕ', either the quantifier depth has gone down or the number of quantifiers at the maximum depth. Therefore by induction hypothesis, $\text{L}(\phi')$ is recognized by $M \in \Box^*_w U_1$. Note that $\text{L}(\phi) = \sigma^{-1}(\text{L}(\phi'))$. By Proposition 5, we get $\text{L}(\phi)$ is recognized by $\wp(\Sigma)\Box M$ which by Lemma 12 is a weakly iterated block product of U_1s. $\Box$

6. First Order Logic with Infinitary Quantifiers - FO[∞]

We now move on to characterizing higher classes of logics like first order logic. In the classical setting, FO has a nice block product based decompositional characterization (see [15]). Our next theorem (Theorem 7) shows that a similar characterization holds for FO interpreted over countable words. Next we introduce an extended version of first order logic, namely FO[∞], that admits nice decompositional characterization using block products. In fact, the characterization results for FO[∞] subsume those for FO and its single variable fragment. In this section, our aim is to introduce this new logic, explain its motivation, and also place it in terms of well studied logics over countable words. We first provide block product characterization of $\oplus$-semigroups recognizing FO languages over linear countable orderings.

Theorem 7. *A language over countable words is definable in* FO *if and only if it is recognized by an iterated block product of* U_1*s.*

We skip the proof here since this theorem can be seen as a corollary of Theorem 10 in the next section.

Our results for FO and its syntactic fragments (see Theorem 3, Theorem 4, Theorem 6 and Theorem 7) closely resemble the corresponding results over finite words. This can be attributed to the limited capability of the operators τ, τ^* and κ in the syntactic $\oplus$-algebra corresponding to FO languages. For instance, FO cannot define the language of words with infinite number of a's [13] — a natural property in the context of countable words. The existential quantifier of FO is a threshold counting quantifier; it says there exists at least one position satisfying a property. Using multiple such first-order quantifiers, FO can count up to any finite constant but not more. Over countable words, it is natural to ask for stronger threshold quantifiers. We introduce natural infinitary extensions of the existential quantifier.

Let $\mathcal{I}_0$ be the set of all non-empty finite orderings. For any number $n \in \mathbb{N}$, we define the set $\mathcal{I}_n$ to be the set of all non-empty orderings of the form $\sum_{i\in\mathbb{Z}} \alpha_i$ where $\alpha_i \in \mathcal{I}_{n-1} \cup \{\varepsilon\}$ and is closed under finite sum. We define the *infinitary rank* (or simply *rank*) of a linear ordering α (denoted by ∞-*rank*(α)) as the least n (if it exists) where $\alpha \in \mathcal{I}_n$. If there is no such n we say that the rank is infinite. For example, $\infty\text{-}rank(\omega) = \infty\text{-}rank(\omega+\omega) = \infty\text{-}rank(\omega^*+\omega) = 1$, $\infty\text{-}rank(\omega^2) = \infty\text{-}rank(\omega^2+\omega^*) = 2$, and the rank of $\eta = (\mathbb{Q}, <)$ is infinite.

We introduce the logic FO$[\infty]$ extending FO with infinitary quantifiers:

$$\varphi := a(x) \mid x < y \mid \varphi \vee \varphi \mid \neg\varphi \mid \exists x\ \varphi \mid \exists^{\infty_0} x\ \varphi \mid \ldots \mid \exists^{\infty_n} x\ \varphi \mid \ldots \qquad n \in \mathbb{N}$$

Note that all the variables are first order and they are interpreted as positions, that is, elements of the underlying linear ordering. More precisely, models of FO$[\infty]$ formula are of the form $w, \mathcal{A}$ where w is a countable word over Σ and $\mathcal{A}$ is an assignment of free (or unquantified) variables to positions in w. The semantics of the new infinitary quantifier $\exists^{\infty_n} x$ is: for a word w and an assignment $\mathcal{A}$, we say $w, \mathcal{A} \models \exists^{\infty_n} x\ \varphi$ if there exists a subordering $X \subseteq \mathrm{dom}(w)$ such that $\infty\text{-}rank(X) = n$ and $w, \mathcal{A}[x=i] \models \varphi$ for all $i \in X$. Here $\mathcal{A}[x=i]$ denotes an assignment $\mathcal{A}'$ which is defined as: $\mathcal{A}'(x) = i$ and $\mathcal{A}'(y) = \mathcal{A}(y)$ for all $y \neq x$. For example, $\exists^{\infty_0} x\ \varphi$ is equivalent to $\exists x\ \varphi$ since both formulas are true if and only if there is at least one satisfying assignment for x. The rest of the semantics is standard.

The logic FO$[(\infty_j)_{j\leq n}]$ denotes the fragment containing only the infinitary quantifiers $\exists^{\infty_j} x$ for all $j \leq n$. Clearly the following natural hierarchy is maintained among the logics:

$$\mathrm{FO} = \mathrm{FO}[(\infty_j)_{j\leq 0}] \subseteq \mathrm{FO}[(\infty_j)_{j\leq 1}] \subseteq \mathrm{FO}[(\infty_j)_{j\leq 2}] \subseteq \ldots$$

We also denote by $\mathrm{FO}^1[(\infty_j)_{j\leq n}]$ the corresponding one variable fragment of $\mathrm{FO}[(\infty_j)_{j\leq n}]$.

Example 12. The formula $\exists^{\infty_1} x\ a(x)$ denotes the set of all countable words with infinitely many positions labelled a. Since FO cannot express this, it shows $\mathrm{FO} \subsetneq \mathrm{FO}[(\infty_j)_{j\leq 1}]$.

Example 13. Consider the language L of all words with $a^\omega a^{\omega^*}$ as a factor. Suppose we have a formula $\mathtt{inf}(x, y)$ that can express that there are infinitely many positions between x and y (assuming $x < y$). We define L using this formula as follows. Guess two 'endpoints' x and y of the factor $a^\omega a^{\omega^*}$. We express the following properties for the positions in this non-empty interval: (1) every position is labelled a, (2) every position is finite distance away from one endpoint and infinite distance away from the other, (3) the points that are finite distance away from the left endpoint have no maximum, and (4) the points that are finite distance away from the right endpoint have no minimum.

1. $\psi_1(x, y) ::= \forall z\ x \leq z \leq y \Rightarrow a(z)$
2. $\psi_2(x, y) ::= \forall z\ x \leq z \leq y \Rightarrow (\neg\mathtt{inf}(x, z) \wedge \mathtt{inf}(z, y)) \vee (\mathtt{inf}(x, z) \wedge \neg\mathtt{inf}(z, y))$
3. $\psi_3(x, y) ::= \forall z\ (x < z < y \wedge \neg\mathtt{inf}(x, z)) \Rightarrow \exists z'\ z < z' < y \wedge \neg\mathtt{inf}(x, z')$
4. $\psi_4(x, y) ::= \forall z\ (x < z < y \wedge \neg\mathtt{inf}(z, y)) \Rightarrow \exists z'\ x < z' < z \wedge \neg\mathtt{inf}(z', y)$

The sentence $\exists x \exists y\ x < y \wedge \psi_1(x, y) \wedge \psi_2(x, y) \wedge \psi_3(x, y) \wedge \psi_4(x, y)$ defines the language L. It is easy to check that $\exists^{\infty_1} z\ x < z < y$ expresses the property $\mathtt{inf}(x, y)$. Therefore L is $\mathrm{FO}[\infty]$ definable.

We now place the logic $\mathrm{FO}[\infty]$ amidst the logics studied in the context of countable words [10, 19]. The logic FO[cut] is an extension of FO that allows quantification over downward closed sets, also known as Dedekind-cuts. Syntactically, we write $\exists_{cut} X$ to existentially quantify a set X where X is downward closed because of the quantifier. The logic WMSO allows quantification over finite subsets of positions. We write $\forall_{fin} X$ to universally quantify over finite sets; here X is a finite set because of the quantifier.

Example 14. Let α be an ordering which contains an ω sequence of positions $(a_i)_{i\in\mathbb{N}}$. Now consider the set $X = \{x \in \alpha \mid x < a_i \text{ for some } i \in \mathbb{N}\}$. It is clearly a downward closed set and thus defines a cut. Furthermore this set has no maximum position, since for any $x \in X$, if $x < a_i$ then there exists $z \in X$ where $x < a_i < z < a_{i+1}$. Therefore we have shown that any ordering containing an ω sequence of positions contains a right-open cut (that is, the downward closed set corresponding to the cut has no maximum element). Conversely, if an ordering contains a right-open cut, then clearly it has an ω sequence of positions. Therefore the FO[cut] formula $\exists_{cut}X\ \exists x\ X(x) \wedge \forall y\ X(y) \Rightarrow \exists z\ X(z) \wedge y < z$ describes the language of all countable words containing an ω sequence of positions.

Example 15. Recall from Example 13 the formula $\texttt{inf}(x, y)$ that expresses there are infinitely many positions between x and y (assuming $x < y$). It was shown that the language L of all words with $a^\omega a^{\omega^*}$ as a factor is definable if $\texttt{inf}(x, y)$ is definable. Now note that $\texttt{inf}(x, y)$ can be defined in WMSO as $\forall_{fin}X\ \exists z\ x < z < y \wedge \neg X(z)$. Therefore L is WMSO definable. It is also possible to define $\texttt{inf}(x, y)$ in FO[cut] because if there are infinitely many positions between x and y then there must be an ω sequence or an ω^* sequence of positions in this interval, and FO[cut] can guess an appropriate cut between x and y to check this. So L is also FO[cut] definable.

In fact, we claim that both first order logic with cuts (FO[cut]) and weak monadic second order logic (WMSO) can define all the languages definable in FO[∞].

Theorem 8. FO[∞] $\subseteq$ FO[cut] $\cap$ WMSO [2]

Proof. We first show by structural induction that there is an equivalent WMSO formula for any FO[∞] formula. It is easy to observe that the hypothesis holds for the atomic case, first order quantification and boolean combinations. Let us consider the formula $\phi = \exists^{\infty_k} x\ \psi(x)$. By our inductive hypothesis there is a WMSO formula $\hat{\psi}(x)$ equivalent to $\psi(x)$. We show that the WMSO formula Ψ_k inductively defined is equivalent to ϕ: Let $\Psi_0 ::= \exists x\ \hat{\psi}(x)$ and

$\Psi_n ::=$ For any finite set $X = \{x_1, \ldots, x_k\}$, one of the factors $[-, x_1], \ldots,$ $[x_i, x_{i+1}], \ldots, [x_k, -]$ can be split into at least two parts each satisfying Ψ_{n-1}

[2]Here, FO[∞], FO[cut], WMSO denote the languages defined by the respective logic.

This can be expressed in WMSO. Note $\texttt{notempty}(X) = \exists x\ X(x)$ says that X is not empty set. Let $\texttt{consec}(X, x, y)$ express that $x, y \in X$ and $x < y$ and there is no $z \in X$ such that $x < z < y$; that is x and y are consecutive in set X. Let $\texttt{min}(X, x)$ denote that x is the minimum position in X, and $\texttt{max}(X, x)$ denote that x is the maximum position in X. Then we define Ψ_n to be

$$\begin{aligned}
&\forall_{fin} X\ \Big(\texttt{notempty}(X) \Rightarrow \\
&\exists x, y, z\ \texttt{consec}(X, x, y) \wedge x < z < y \wedge \Psi_{n-1}[> x, < z] \wedge \Psi_{n-1}[> z, < y]) \vee \\
&\exists x, z\ \big(\texttt{min}(X, x) \wedge z < x \wedge \Psi_{n-1}[> z, < x] \wedge \Psi_{n-1}[< z]\big)\ \vee \\
&\exists x, z\ \big(\texttt{max}(X, x) \wedge x < z \wedge \Psi_{n-1}[> x, < z] \wedge \Psi_{n-1}[> z]\big)\Big)
\end{aligned}$$

We claim that Ψ_n is satisfied by all words where the ψ-labelled set of positions α has ∞-*rank*$(\alpha) \geq n$. It is clearly true for the base case Ψ_0. Assume the hypothesis is true for all $j < n$. The formula Ψ_n says that for any finite number of partitions $\alpha_1, \alpha_2, \ldots, \alpha_k$, of the ψ-labelled set of positions α, there is at least one α_i that can be split into two parts containing ψ-labelled set of positions α_i^1 and α_i^2 such that ∞-*rank*$(\alpha_i^1) \geq n-1$ and ∞-*rank*$(\alpha_i^2) \geq n-1$. In short, finite partitioning of ψ-labelled set of positions with rank $n-1$ is not possible or ∞-*rank*$(\alpha) \geq n$. Therefore the formula Ψ_k is equivalent to the formula ϕ.

Next we give an FO[cut] formula equivalent to an FO[∞] formula. Like in the previous proof, let us look at the case $\phi = \exists^{\infty_k} x\ \psi(x)$ where $\psi(x)$ is equivalent to an FO[cut] formula $\hat{\psi}(x)$. We show ϕ is equivalent to Φ_k where Φ_n is inductively defined as: $\Phi_0 ::= \exists x\ \hat{\psi}(x)$ and Φ_n is

> There is a cut towards which there is an ω (or ω^*) sequence of factors each satisfying Φ_{n-1}

Let X be a non-empty cut. We give an FO[cut] formula $\texttt{omegaseq}(X)$ that says there is an ω sequence of factors satisfying Φ_{n-1} approaching towards the cut X.

$$\texttt{omegaseq}(X) ::= \forall y\ X(y) \Rightarrow \exists z\ X(z) \wedge y < z \wedge\ \Phi_{n-1}[> y, < z]$$

The formula says there is an ω sequence of positions such that each factor between consecutive positions contains ψ-labelled subsequence of rank $\geq n-1$. Similarly, there is a formula $\texttt{omegaseq}^*(X)$ that state the existence of an ω^* sequence approaching the cut. The formula Φ_n will guess this cut and verify the ω or ω^* sequence is non-empty as given below.

$$\Phi_n ::= \exists_{cut} X \ \Big(\exists x \ X(x) \wedge \texttt{omegaseq}(X)\Big) \ \vee \ \Big(\exists x \ \neg X(x) \wedge \texttt{omegaseq}^*(X)\Big)$$

Inductively arguing about the correctness of the formula, it's quite clear that Φ_n expresses existence of set of ψ-labelled positions of rank $\geq n$. □

7. Product Decompositions for FO[∞]

We now apply our algebraic tools to give decompositional characterizations of FO[∞] and its one variable fragments. Our approach uses the block product principle that we developed in subsection 4.4 to directly show equivalence of languages definable in some logic and languages recognized by some family of $\oplus$-semigroups.

We first identify a family of simple $\circledast$-algebras that will help characterize the logic. For $n \geq 0$, let $\Delta_n = (\{1, \delta_0, \delta_1, \ldots, \delta_n\}, \cdot, \tau, \tau^*, \kappa)$ be an $\circledast$-algebra where

- $\delta_i \cdot \delta_j = \delta_j \cdot \delta_i = \delta_j$ for all $0 \leq i \leq j \leq n$
- ${\delta_k}^{\tau} = {\delta_k}^{\tau^*} = \delta_{k+1}$ for all $0 \leq k < n$, and ${\delta_n}^{\tau} = {\delta_n}^{\tau^*} = \delta_n$
- $S^{\kappa} = \delta_n$ for all $S \backslash \{1\} \neq \emptyset$

It is easy to verify that Δ_n is an idempotent and commutative $\circledast$-algebra. Further, observe that Δ_n is generated by the element δ_0.

7.1. FO[∞] *with single variable*

In this subsection we show that languages recognized by Δ_n are definable in $\mathrm{FO}^1[(\infty_j)_{j\leq n}]$. It easily follows that the direct product of Δ_n recognize exactly those languages definable in the one variable fragment, which is our next theorem.

Theorem 9. *Languages recognized by direct product of Δ_n are exactly those definable in* $\mathrm{FO}^1[(\infty_j)_{j\leq n}]$.

Proof. We first show that languages recognized by Δ_n are those definable in $\mathrm{FO}^1[(\infty_j)_{j\leq n}]$. In this proof, we adopt the convention that $1 = \delta_{-1}$. Let $h \colon \Sigma^{\oplus} \to \Delta_n$ be a morphism. It suffices to show that for any element

$\delta_m \in \Delta_n$, $h^{-1}(\delta_m)$ is definable in $\mathrm{FO}^1[(\infty_j)_{j \le n}]$. Let $\uparrow m$ denote the set $\{\delta_{m'} \mid m' \ge m\}$. Note that for any $\delta_m \ne \delta_n$, $h^{-1}(\delta_m) = h^{-1}(\uparrow m) \setminus h^{-1}(\uparrow(m+1))$. Also $h^{-1}(\delta_n) = h^{-1}(\uparrow n)$. Therefore, it is sufficient to show that $h^{-1}(\uparrow m)$ is definable in $\mathrm{FO}^1[(\infty_j)_{j \le n}]$.

For each $m = \{-1, 0, \ldots, n\}$, we define the language $L(m)$ as the set of all words with at least one of the following two properties

- there exists a letter a in w such that $h(a) \in \uparrow m$
- there is a nonempty subordering $\alpha \subseteq \mathrm{dom}(w)$ whose all positions are labelled a, the ∞-rank of α is j, $h(a) = \delta_i \ne \delta_{-1}$ and $i + j \ge m$

The following $\mathrm{FO}^1[(\infty_j)_{j \le n}]$ sentence defines the language $L(m)$.

$$\bigvee_{a \in \Sigma,\ h(a) \in \uparrow m} \exists x\ a(x) \quad \vee \quad \bigvee_{\substack{a \in \Sigma,\ h(a) = \delta_i \ne 1 \\ i + j \ge m}} \exists^{\infty_j} x\ a(x)$$

We show that $L(m) = h^{-1}(\uparrow m)$ by induction on the m. For $m = -1$, this clearly holds as $\uparrow\{-1\} = \Delta_n$, and therefore $h^{-1}(\uparrow\{-1\}) = \Sigma^{\oplus}$, and also $L(-1) = \Sigma^{\oplus}$. To prove the induction hypothesis assume the claim holds for all $m' < m$. Consider a word w. By a second induction on the height of an evaluation tree (T, h) for w we show for all words $v \in T$, $v \in h^{-1}(\uparrow m)$ if and only if $v \in L(m)$. In each of the following cases we assume that the children of the node (if they exist) satisfy the second induction hypothesis.

1. Case v is a letter: The hypothesis clearly holds
2. Case v is a concatenation of two words v_1 and v_2: There are two cases to consider - $\{v_1, v_2\} \cap h^{-1}(\uparrow m) \ne \emptyset$ or not. In the first case, let for an $i \in \{1, 2\}$ we have $h(v_i) \in \uparrow m$ and $v_i \in L(m)$. Clearly $h(v) = h(v_1 v_2) \in \uparrow m$ and $v \in L(m)$. For the second case, let us assume $h(v_1) = \delta_{m_1}$ and $h(v_2) = \delta_{m_2}$ such that $m_1 \le m_2 < m$ and both $v_1, v_2 \notin L(m)$. From the definition of Δ_n, it follows that $h(v) = h(v_1 v_2) = \delta_{m_2}$. For any $a \in \Sigma$, let the a-labelled suborderings in v_1 and v_2 be α_1 and α_2 respectively where $\infty\text{-}\mathit{rank}(\alpha_1) \le \infty\text{-}\mathit{rank}(\alpha_2) = j$. It follows from the definition that $\infty\text{-}\mathit{rank}(\alpha_1 + \alpha_2) = j$ and therefore $v \notin L(m)$.
3. Case v is an ω-sequence of words $\langle v_1, v_2, \ldots, \rangle$ such that $h(v_i) = \delta_{m'}$ for all i, and $\delta_{m'}$ is an idempotent (in Δ_n all elements are idempotents):

Firstly, if $m' \geq m$ and $v_i \in L(m)$ then clearly $h(v) \in \uparrow m$ and $v \in L(m)$. The non-trivial case is $m' = m - 1$. From the second induction hypothesis $v_i \notin L(m)$ for all i. If $\delta_{m'} = 1$, then $h(v) = 1 \notin \downarrow m$ and $v \notin L(m)$. Otherwise from the definition of Δ_n, $h(v) = (\delta_{m'})^\tau = \delta_m$, and each factor v_i contains some letter mapping to non-neutral elements of Δ_n. We need to show that $v \in L(m)$. By first induction hypothesis, each v_i has a letter a_i and an a_i-labelled set of positions α_i such that $h(a_i) = \delta_{k_i}$ and ∞-*rank*$(\alpha_i) = k'_i$ such that $k_i + k'_i \geq m'$. Since $|\Sigma|$ is finite, ω-many of these a_is are the same letter, say a. Let $h(a) = \delta_k$. Then for all i such that $a_i = a$, we know ∞-*rank*$(\alpha_i) \geq k'$ where $k + k' \geq m'$. Hence the a-labelled set of positions $\alpha = \sum_{i:a_i=a} \alpha_i$ in v satisfies ∞-*rank*$(\alpha) \geq k'+1$, and since $k+k'+1 \geq m$ we get $v \in L(m)$.

4. Case v is an ω^*-sequence: This case is symmetric to the above case.

5. Case v is $\prod_{i\in\eta} v_i$, $\prod_{i\in\eta} h(v_i)$ is a perfect shuffle of $\{h(v_i) | i \in \eta\} = S$ and $h(v) = S^\kappa$: It is easy to see that the induction hypothesis holds if $S = \{1\}$. So, assume $S \backslash \{1\} \neq \emptyset$. Hence $h(v) = \delta_n$. Since, there are η-many of children u where $h(u) \neq 1$, there is a letter a such that $h(a) \neq 1$ and a-labelled set of positions in v has infinite ∞-rank. Thus $v \in L(n)$.

For the other direction, note that Δ_n recognizes the language $\exists^{\infty_i} x\ (a(x) \vee b(x))$ for $i \leq n$ by the morphism $h(a) = h(b) = \delta_{n-i}$ and for $c \notin \{a, b\}$, $h(c) = 1$; the language then is $h^{-1}(\delta_n)$. The proof follows from the fact that a one variable quantifier free formula is essentially a disjunction of letter predicates and therefore the boolean combination of sentences can be recognized by direct products of Δ_n. □

We now provide an equational algebraic characterization of the syntactic ⊛-algebras of languages definable in $\mathrm{FO}^1[(\infty_j)_{j \leq n}]$. This is achieved by formulating an equational description of algebras which divide direct product of Δ_n.

We begin with the definition of a *shuffle-n-symmetric-trivial* algebra. We say that a ⊕-algebra $(M, \cdot, \tau, \tau^*, \kappa)$ is shuffle-n-symmetric-trivial if M satisfies the following identities: 1) $x \cdot x = x$ – every element of M is idempotent, 2) $x \cdot y = y \cdot x$ – M is commutative, 3) $x^\tau = x^{\tau^*}, (xy)^\tau = x^\tau y^\tau$, and 4) $x_1^{\tau^n} \cdot x_2^{\tau^n} \cdot \ldots \cdot x_p^{\tau^n} = \{x_1, \ldots, x_p\}^\kappa$ where $x^{\tau^0} = x$ and $x^{\tau^{i+1}} = \left(x^{\tau^i}\right)^\tau$. Note that the definition of 'shuffle-trivial' from subsection 3.1 matches that of shuffle-n-symmetric-trivial when n is 0.

Proposition 1. *Let M be a finite $\circledast$-algebra. Then M divides a direct product of Δ_n iff M is shuffle-n-symmetric-trivial.*

Proof. It is clear that Δ_n is shuffle-n-symmetrical trivial and this property is preserved under direct product and division. This shows that if M divides a direct product of Δ_n then it is shuffle-n-symmetric-trivial.

For the converse, we fix a shuffle-n-symmetric-trivial M. It is easy to deduce that, for any element m of M, the subalgebra $\langle m \rangle$ of M generated by m is isomorphic to Δ_k for some $k \leq n$. In fact, the underlying set of $\langle m \rangle$ consists of elements $\{1, m = m^2, m^\tau = m^{\tau^*}, \ldots, m^{\tau^k} = m^{\tau^{k+1}} = m^\kappa\}$ and the well-defined morphism obtained by sending the generator of Δ_k to m provides an isomorphism between Δ_k and $\langle m \rangle$. We also have a morphism h_m from Δ_n to M which maps the generator of Δ_n to m such that the image of h_m is precisely $\langle m \rangle$.

Let $S = \{m_1, m_2, \ldots m_p\}$ be a generating set of M. An important consequence of shuffle-n-symmetric-triviality of M is that every element of M can be expressed as $m_1^{\tau^{i_1}} m_2^{\tau^{i_2}} \cdots m_p^{\tau^{i_p}}$ where $0 \leq i_1, i_2, \ldots, i_p \leq n$.

We can now construct a map $h : \prod_1^p \Delta_n \to M$ by combining the individual morphisms $h_{m_i} : \Delta_n \to M$ as follows:

$$h((n_1, n_2, \ldots, n_p)) = h_{m_1}(n_1) h_{m_2}(n_2) \cdots h_{m_p}(n_p)$$

It can be argued that h is a surjective morphism. We skip the straightforward details. This shows that M is a homomorphic image of a direct product of Δ_n and completes the proof. □

Combining the above proposition with Theorem 9, we conclude that a language is definable in $\mathrm{FO}^1[(\infty_j)_{j \leq n}]$ iff its syntactic $\circledast$-algebra is shuffle-n-symmetric trivial. Thus we also obtain a decidable equational algebraic characterization of the one variable fragment $\mathrm{FO}^1[(\infty_j)_{j \leq n}]$.

7.2. Block Product Decompositions for FO$[\infty]$

In this section, we consider the full logic $\mathrm{FO}[(\infty_j)_{j \leq n}]$ and observe that they define exactly those languages recognized by block products of Δ_n. First we show relativizing $\mathrm{FO}[(\infty_j)_{j \leq n}]$ formulas with respect to first order variables works as intended. We'll only use this result for $\mathrm{FO}[(\infty_j)_{j \leq n}]$ sentences though. See [15, Lemma VI.1.3] for a similar proof for FO.

Lemma 13. *Let $\varphi \in \mathrm{FO}[(\infty_j)_{j \leq n}]$ be a formula. Consider any word w with an assignment $\mathcal{A}$ that maps elements of $\mathrm{free}(\varphi)$ to positions less than some position $i \in \mathrm{dom}(w)$. If $x \notin \mathrm{free}(\varphi)$, then we can construct a relativized formula $\varphi_{<x}$ with $\mathrm{free}(\varphi_{<x}) = \mathrm{free}(\varphi) \cup \{x\}$ such that*

$$w, \mathcal{A}[x = i] \models \varphi_{<x} \textit{ iff } w_{<i}, \mathcal{A} \models \varphi$$

Proof. Proof is via structural induction on $\mathrm{FO}[(\infty_j)_{j \leq n}]$ formula. We only show the case for the extended infinitary quantifier. Let $\varphi = \exists^{\infty_k} y\ \psi$. We note that $w_{<i}, \mathcal{A} \models \exists^{\infty_k} y\ \psi$ if and only if there is a subordering $X \subseteq \mathrm{dom}(w_{<i})$ such that $\infty\text{-}rank(X) = k$ and for all $j \in X$, $w_{<i}, \mathcal{A}[y = j] \models \psi$. It follows, from the inductive hypothesis, that this is true if and only if $w, \mathcal{A}[x = i] \models \exists^{\infty_k} y (\psi_{<x} \wedge y < x)$. This completes the proof. □

Theorem 10. *The languages defined by $\mathrm{FO}[(\infty_j)_{j \leq n}]$ are exactly those recognized by finite block products of Δ_n. Moreover, the languages defined by $\mathrm{FO}[\infty]$ are exactly those recognized by finite block products of $\{\Delta_n \mid n \in \mathbb{N}\}$.*

Proof. We first show that languages recognizable by finite block products of Δ_n are definable in $\mathrm{FO}[(\infty_j)_{j \leq n}]$. The proof is via induction on the number of Δ_n in an iterated block product. The base case follows from Theorem 9.

For the inductive step, consider a morphism $h \colon \Sigma^\oplus \to M \Box \Delta_n$. Let $h_{\mathbb{1}} \colon \Sigma^\oplus \to M$ be the induced morphism to M, and let σ be the associated transducer. By the block product principle (see Proposition 5), any language recognized by h is a boolean combination of languages $L_1 \subseteq \Sigma^\oplus$ recognized by M and $\sigma^{-1}(L_2)$ where $L_2 \subseteq (M \times \Sigma \times M)^\oplus$ is recognized by Δ_n. By induction hypothesis, L_1 is $\mathrm{FO}[(\infty_j)_{j \leq n}]$ definable. By the base case L_2 is $\mathrm{FO}[(\infty_j)_{j \leq n}]$ definable but over the alphabet $M \times \Sigma \times M$. To complete the proof, one needs to show for any word $w \in \Sigma^\oplus$ and assignment s, and for any $\mathrm{FO}[(\infty_j)_{j \leq n}]$ formula φ over the alphabet $M \times \Sigma \times M$, there exists a $\mathrm{FO}[(\infty_j)_{j \leq n}]$ formula $\hat{\varphi}$ over the alphabet Σ such that $w, s \models \hat{\varphi}$ if and only if $\sigma(w), s \models \varphi$. For instance, suppose $\varphi = \exists^{\infty_i} x\ (m_1, c, m_2)(x)$, and inductively ϕ_{m_1} (resp. ϕ_{m_2}) are $\mathrm{FO}[(\infty_j)_{j \leq n}]$ sentences characterizing words over $\Sigma^\oplus$ that are mapped by $h_{\mathbb{1}}$ to m_1 (resp. m_2). Then $\hat{\varphi}$ is $\exists^{\infty_i} x\ ((\phi_{m_1})_{<x} \wedge c(x) \wedge (\phi_{m_2})_{>x})$, where $(\phi_{m_1})_{<x}$ is the formula ϕ_{m_1} relativized to less than the variable x. This way, one proves that $\sigma^{-1}(L_2)$ is $\mathrm{FO}[(\infty_j)_{j \leq n}]$ definable. This completes the proof of this direction.

The other direction of the proof is a standard generalization of the proof for FO in the classical setting [15]. It progresses via structural induction on

FO$[(\infty_j)_{j\leq n}]$ formulas. We know that FO$[\infty]$ has letter and order predicates, is closed under boolean operations and infinitary existential quantifications. Inductively we prove that for any FO formula $\varphi = \phi(x_1, x_2, \ldots, x_n)$, the language $\mathrm{L}(\varphi) \subseteq (\Sigma \times \{0,1\}^n)^\oplus$ over extended alphabet is recognized by an iterated block product of U_1. In this proof, we call a word/model valid if the 'row' for each variable contains exactly one position labelled 1.

For the base case, let $\varphi = a(x)$. The language of this formula is the set of all valid words with an occurence of $(a, 1)$ (validity of the word enforces exactly one occurence of $(a, 1)$). Recalling Example 11 one can see that checking validity of words can be done by direct product of copies of $U_1\Box U_1$. In particular, the language for $a(x)$ can be recognized by $U_1 \times (U_1\Box U_1)$ (also recall Example 5), and by Lemma 11, this divides an iterated block product of U_1s. Similarly, it is easy to show that language defined by $x < y$ is recognized by iterated block products of U_1. Boolean combinations of first order formulas can be inductively recognized by direct product of the algebras for individual formulas (extra validity checks, if required, for instance, for negation, can be handled as per our discussion so far). The non-trivial case is when $\phi = \exists^{\infty_i} x\ \psi$ (for $i \leq n$). Let $L(\psi) \subseteq (\Sigma \times \{0,1\})^\oplus$ be inductively recognized by $h\colon (\Sigma \times \{0,1\})^\oplus \to M \in \Box^*\Delta_n$, that is, there is a set $F \subseteq M$ such that $h^{-1}(F) = L(\psi)$. We prove that $M\Box\Delta_n$ recognizes $L(\phi)$. Once again we use the block product principle. Consider two morphisms $g_1\colon \Sigma^\oplus \to M$ and $g_2\colon (M \times \Sigma \times M)^\oplus \to \Delta_n$. Let $g_1(a) = h((a, 0))$ and suppose $g_2((m_1, a, m_2))$ equals δ_0 if $m_1{\cdot}h((a, 1)){\cdot}m_2 \in F$, and it equals 1 otherwise. Let σ be the transducer corresponding to g_1. We show that $w \models \phi$ if and only if $g_2(\sigma(w)) = \delta_j$ where $j \geq i$. This would imply $L(\phi) = \sigma^{-1}(g_2^{-1}(\{\delta_i, \delta_{i+1}, \ldots, \delta_n\}))$ and by the block product principle, this is recognized by $M\Box\Delta_n$.

Let $w \models \phi$. If α_ψ is the set of all positions of w where ψ is true, then ∞-$rank(\alpha_\psi) \geq i$. Let $l \in \alpha_\psi$ and $w(l) = a$. We can split w at the position l as $w_1 a w_2$ and by logic semantics $w_1^0(a, 1)w_2^0 \models \psi$ (for any $u \in \Sigma^\oplus$, we denote by u^0 the word over the same domain with $u^0[i] = (u[i], 0)$). If $h(w_1^0) = m_1$ and $h(w_2^0) = m_2$, then $m_1 \cdot h((a, 1)) \cdot m_2 \in F$. Also, $\sigma(w)[l] = (m_1, a, m_2)$. So, g_2 maps every position $l \in \alpha_\psi$ to δ_0, and hence $g_2(\sigma(w)) = \delta_j$ for some $j \geq i$. Conversely, suppose $g_2(\sigma(w)) = \delta_j$ where $j \geq i$. Let α_0 denote the positions of $\sigma(w)$ for which g_2 maps to δ_0. Since g_2 maps each letter to δ_0 or 1, we get ∞-$rank(\alpha_0) \geq i$. Let $l \in \alpha_0$. If $\sigma(w)[l] = (m_1, a, m_2)$, then $m_1 \cdot h((a, 1)) \cdot m_2 \in F$. This means ψ is true at position l for w. Since l is any position in α_0, we have that $w \models \phi$. $\Box$

8. No Finite Block Product Basis Results

The main goal of this section is to prove that $\mathrm{FO}[\infty], \mathrm{FO}[\mathrm{cut}]$, and the semantic class $\mathrm{FO}[\mathrm{cut}] \cap \mathrm{WMSO}$ over countable words do not admit a block product based characterization which uses only a *finite* set of $\oplus$-algebras (Theorem 12). This is achieved by defining a suitable parameter called *gap-nesting-length* for $\oplus$-algebras (Definition 6), and our main technical lemma of this section, Lemma 18, that shows the parameter value does not increase on division and block product (for block product, we assume aperiodicity). This lemma also establishes that the infinite syntactic hierarchy inside $\mathrm{FO}[\infty]$ to be strict (Theorem 11).

The result of Theorem 12 is in stark contrast to our previous result over FO, Theorem 7 which shows that a language of countable words is FO-definable if and only if it is recognized by a strong iteration of block product of copies of the single $\circledast$-algebra U_1 (alternately Δ_0). In the last section Theorem 10 shows that $\mathrm{FO}[\infty]$ has a block product characterization using the natural infinite basis set $\{\Delta_n\}_{n\in\mathbb{N}}$. The results in this section prove that this is optimal.

Fix a finite $\oplus$-algebra $(M, \cdot, \tau, \tau^*, \kappa)$. For every $n \in \mathbb{N}$, we define the operation $\gamma_n : M \to M$ which maps x to x^{γ_n}. The inductive definition of γ_n is as follows (recall that idempotent power is denoted by !): $x^{\gamma_0} = x^!$ and $x^{\gamma_n} = ((x^{\gamma_{n-1}})^{\tau}(x^{\gamma_{n-1}})^{\tau^*})^!$.

Lemma 14. *Let M be a finite $\oplus$-algebra. For each $m \in M$, there exists $n \in \mathbb{N}$ such that $\forall n' \geq n, m^{\gamma_n} = m^{\gamma_{n'}}$.*

Proof. Consider the following sequence: $a_0 = m^!$ and $a_{j+1} = ((a_j)^{\tau} \cdot (a_j)^{\tau^*})^!$. Clearly, $a_i = m^{\gamma_i}$; we prove this sequence becomes constant beyond a finite index. By $\oplus$-algebra axioms $x \cdot x^{\tau} = x^{\tau}$ and $x^{\tau^*} \cdot x = x^{\tau^*}$, we get that $a_{j+1} = a_j \cdot a_{j+1} = a_{j+1} \cdot a_j$ for all j. This and the fact that every element of this sequence is an idempotent further implies that for all $i \leq j$, we have $a_j = a_i \cdot a_{i+1} \dots a_j$.

Since M is finite, there is an i and a $j > i$ such that $a_i = a_j$. Let us assume that j is the smallest index strictly larger than i such that $a_i = a_j$. It is sufficient to show that $j = i+1$. We know $a_j = a_j \cdot a_{j-1}$. Since $a_i = a_j$, we get that $a_i = a_i \cdot a_{j-1}$. As $i \leq j-1$, we also know that $a_{j-1} = a_i \cdot a_{i+1} \dots a_{j-1}$. Therefore,

$$a_i = a_i \cdot a_{j-1} = a_i \cdot a_i \cdot a_{i+1} \dots a_{j-1} = a_i \cdot a_{i+1} \dots a_{j-1} = a_{j-1}$$

By the minimality of j, we get that $j-1=i$, that is, $j=i+1$. □

Definition 6. The *gap-nesting-length* of a $\oplus$-algebra M, denoted $\text{gnlen}(M)$, is the smallest n such that for all $m \in M$, $m^{\gamma_n} = m^{\gamma_{n+1}}$.

It follows from the previous lemma that a finite $\oplus$-algebra has a finite gap-nesting-length. It is a simple computation that, for each k, $\text{gnlen}(\Delta_k) = k$. The main technical lemma of this section is Lemma 18 that states that the gap-nesting-length parameter does not increase on division and block product of $\oplus$-algebras. This is the key to our no-finite-basis theorems. The following couple of results will help us prove the main lemma.

Lemma 15. *Consider $\oplus$-algebra M has compatible left and right actions on $\oplus$-algebra P. Let $m, m' \in M$ and $p \in P$. Then $mp^{\gamma_n}m' = (mpm')^{\gamma_n}$*

Proof. We first prove that $mp^! m' = (mpm')^!$. By action axioms (recall B-2 for left action), it is easy to see that $mp^k m' = (mpm')^k$ for any natural number $k \geq 1$. Note that there exists $k \in \mathbb{N}$ such that $p^k = p^!$ and $(mpm')^k = (mpm')^!$. Then $mp^! m' = mp^k m' = (mpm')^k = (mpm')^!$.

The proof is now by induction on n. For $n=0$, we have $mp^{\gamma_0}m = mp^! m = (mpm)^! = (mpm)^{\gamma_0}$.

For the inductive step, note that

$$\begin{aligned}
mp^{\gamma_n}m' &= m((p^{\gamma_{n-1}})^{\tau} \cdot (p^{\gamma_{n-1}})^{\tau^*})^! m' && \text{defn. of } \gamma_n \\
&= (m((p^{\gamma_{n-1}})^{\tau} \cdot (p^{\gamma_{n-1}})^{\tau^*})m')^! \\
&= ((m(p^{\gamma_{n-1}})^{\tau} m') \cdot (m(p^{\gamma_{n-1}})^{\tau^*} m'))^! && \text{action axiom for } \cdot \\
&= ((m(p^{\gamma_{n-1}})m')^{\tau} \cdot (m(p^{\gamma_{n-1}})m')^{\tau^*})^! && \text{action axiom for } \tau,\, \tau^* \\
&= (((mpm')^{\gamma_{n-1}})^{\tau} \cdot ((mpm')^{\gamma_{n-1}})^{\tau^*})^! && \text{induction hypothesis} \\
&= ((mpm')^{\gamma_n} && \text{defn. of } \gamma_n
\end{aligned}$$

This completes the proof. □

Lemma 16. *Let M and N be two $\oplus$-algebras where M has compatible actions on N. Let $(m, n), (m', n') \in M \ltimes N$ such that $(m, n) = (m', n')^!$. Then $m = (m')^!$. Further, if M is aperiodic*[3]*, then $mnm = (mn'm)^!$.*

[3] we say a $\oplus$-algebra is aperiodic if its underlying semigroup is aperiodic

Proof. Note that by concatenation rule of semidirect product algebra, we have $(m,n)^2 = (m^2, nm + mn)$. Since (m,n) is an idempotent, we get $m = m^2$, that is, $m \in M$ is an idempotent. Also, we get $n = nm + mn$ which implies $mnm = mnm^2 + m^2nm$. Using the fact that $m = m^2$, we get that mnm is an idempotent in N.

Suppose $k \in \mathbb{N}$ such that of $(m,n) = (m', n')^k$. An easy calculation shows that $m = (m')^k$ and $n = \sum_{i=0}^{k-1} (m')^i n' (m')^{k-i-1}$. By our earlier argument, we know m is an idempotent, so $m = (m')^!$.

If M is aperiodic, then $(m')^j = m$ for $j \geq k$. Hence $mnm = (mn'm)^k$. Since mnm is an idempotent, we get $mnm = (mn'm)^!$. □

Lemma 17. *Consider $(m,f), (m',f') \in M \Box N$ such that $(m,f) = (m',f')^{\gamma_n}$. Then $m = (m')^{\gamma_n}$. If M is aperiodic, then $mfm = (mf'm)^{\gamma_n}$.*

Proof. The proof is by induction on n. For the base case of $n = 0$, we have $(m,f) = (m',f')^{\gamma_0} = (m',f')^!$. By Lemma 16, $m = (m')^! = (m')^{\gamma_0}$ and if M is aperiodic, $mfm = (mf'm)^! = (mf'm)^{\gamma_0}$. This proves the base case.

For the inductive step, let $(m,f) = (m',f')^{\gamma_n} = ((m',f')^{\gamma_{n-1}})^{\gamma_1}$. Also let $(e,g) = (m',f')^{\gamma_{n-1}}$. So $(m,f) = (e,g)^{\gamma_1}$. By induction hypothesis, $e = (m')^{\gamma_{n-1}}$ and $m = e^{\gamma_1}$ implying $m = ((m')^{\gamma_{n-1}})^{\gamma_1} = (m')^{\gamma_n}$. If M is aperiodic, then by induction hypothesis, $ege = (ef'e)^{\gamma_{n-1}}$ and $mfm = (mgm)^{\gamma_1}$. Note that since $m = e^{\gamma_1} = (e^\tau \cdot e^{\tau^*})^!$, we have $m \cdot e = e \cdot m = m$. Therefore

$$\begin{aligned} mfm &= (mgm)^{\gamma_1} \\ &= (m(ege)m)^{\gamma_1} \\ &= (m(ef'e)^{\gamma_{n-1}}m)^{\gamma_1} = ((mf'm)^{\gamma_{n-1}})^{\gamma_1} = (mf'm)^{\gamma_n} \end{aligned}$$

This completes the proof. □

We are now ready to state and prove our main technical lemma of this section.

Lemma 18. *Let M and N be two $\oplus$-algebra.*

1. *If M divides N then* $\mathrm{gnlen}(M) \leq \mathrm{gnlen}(N)$.

2. *If M, N are aperiodic then* $\mathrm{gnlen}(M \Box N) \leq \max(\mathrm{gnlen}(M), \mathrm{gnlen}(N))$.

Proof. 1. If M is a subalgebra of N, then the property is easily verified.

Let's suppose $h \colon N \to M$ is a surjective morphism, and $\mathrm{gnlen}(N) = k$.

For any $m \in M$, there exists $n \in N$ such that $h(n) = m$. It is straightforward to check that $m^{\gamma_k} = h(n^{\gamma_k}) = h(n^{\gamma_{k+1}}) = m^{\gamma_{k+1}}$. This completes the proof for division.

2. Consider aperiodic M and N with $\max(\text{gnlen}(M), \text{gnlen}(N)) = k$. We show that $\text{gnlen}(M \Box N) \leq k$. Note that, for any $m \in M$ and any $n \in N$, $m^{\gamma_k} = m^{\gamma_{k+1}}$ and $n^{\gamma_k} = n^{\gamma_{k+1}}$.

   Let $(m, f) \in M \Box N$ be an arbitrary element. We show that $(m, f)^{\gamma_k} = (m, f)^{\gamma_{k+1}}$. Let $(e, g) = (m, f)^{\gamma_k}$. Then $(e, g)^{\gamma_1} = (m, f)^{\gamma_{k+1}}$. Also by Lemma 17, $e = m^{\gamma_k}$ and $ege = (efe)^{\gamma_k}$. Since M and N have gap-nesting-length less than or equal to k, we get $e = m^{\gamma_k} = m^{\gamma_{k+1}} = e^{\gamma_1}$ and $ege = (efe)^{\gamma_k} = (efe)^{\gamma_{k+1}} = (ege)^{\gamma_1}$. Now we use the fact that in any aperiodic $\oplus$-algebra $x = x^{\gamma_1}$ implies $x = x^{\tau} \cdot x^{\tau^*}$ by the following argument — $x = (x^{\tau} \cdot x^{\tau^*})^! = (x^{\tau} \cdot x^{\tau^*})^! \cdot (x^{\tau} \cdot x^{\tau^*}) = x \cdot (x^{\tau} \cdot x^{\tau^*}) = x^{\tau} \cdot x^{\tau^*}$. Therefore we have $e = e^{\tau} \cdot e^{\tau^*}$ and $ege = (ege)^{\tau} + (ege)^{\tau^*}$. Since (e, g) is an idempotent by definition of the γ_i operation, we get that e is an idempotent in M. Therefore

$$\begin{aligned}
&(e, g)^{\tau} \cdot (e, g)^{\tau^*} \\
&= (e^{\tau} e^{\tau^*}, g e^{\tau} e^{\tau^*} + (e g e^{\tau} e^{\tau^*})^{\tau} + (e^{\tau} e^{\tau^*} g e)^{\tau^*} + e^{\tau} e^{\tau^*} g) \\
&= (e, ge + (ege)^{\tau} + (ege)^{\tau^*} + eg) \\
&= (e, ge + ege + eg) = (e, g)^3 = (e, g)
\end{aligned}$$

   Hence $(m, f)^{\gamma_{k+1}} = (e, g)^{\gamma_1} = (e, g) = (m, f)^{\gamma_k}$. This completes the proof for the block product operation. □

An important application of Lemma 18 is that the syntactic hierarchy inside $\text{FO}[\infty]$ can be shown to be strict.

Theorem 11. $\text{FO}[(\infty_j)_{j \leq n}] \subsetneq \text{FO}[(\infty_j)_{j \leq n+1}]$.

Proof. By Theorem 10, the syntactic $\oplus$-algebra of any $\text{FO}[(\infty_j)_{j \leq n}]$-definable language divides an iterated block product of copies of Δ_n. By Lemma 18 and the fact that $\text{gnlen}(\Delta_k) = k$, $\text{gnlen}(M) \leq n$. Note that, Δ_{n+1} is the syntactic $\circledast$-algebra for the language L defined by the $\text{FO}[(\infty_j)_{j \leq n+1}]$ formula $\exists^{\infty_{n+1}} x\, a(x)$. As $\text{gnlen}(\Delta_{n+1}) = n + 1$, it follows that L cannot be defined in $\text{FO}[(\infty_j)_{j \leq n}]$. □

Finally we present our no-finite-basis theorem.

Theorem 12. *There is no finite basis for a block product based characterization for any of these logical systems* $\mathrm{FO}[\infty], \mathrm{FO}[\mathrm{cut}], \mathrm{FO}[\mathrm{cut}] \cap \mathrm{WMSO}$.

Proof. Fix one of the logics $\mathcal{L}$ mentioned in the statement of the theorem. It follows from Theorem 8 and the decidable algebraic characterization (see [10]) of FO[cut] that the syntactic $\oplus$-algebras of $\mathcal{L}$-definable languages are aperiodic. Now suppose, for contradiction, $\mathcal{L}$ admits a finite basis B of aperiodic $\oplus$-algebras for its block product based characterization. Since B is finite, there exists $n \in \mathbb{N}$ such that for all $\oplus$-algebra M in B, $\mathrm{gnlen}(M) \leq n$. It follows by Lemma 18 that the syntactic $\oplus$-algebra N of *every* $\mathcal{L}$-definable language has the property $\mathrm{gnlen}(N) \leq n$.

Now consider the language L defined by the $\mathrm{FO}[\infty]$ sentence $\exists^{\infty_{n+1}} x\, a(x)$. By Theorem 8, L is $\mathcal{L}$-definable. Hence, the gap-nesting-length of the syntactic $\oplus$-algebra K of L is less than or equal to n. However, K is simply Δ_{n+1} and $\mathrm{gnlen}(\Delta_{n+1}) = n + 1$. This leads to a contradiction. □

9. Conclusion

This work provides various equational as well as product-based decompositional algebraic characterizations of logical formalisms over countable words. Towards this, we have developed a seamless integration of the block product operation into the algebraic framework well suited for the countable setting.

In fact, we have obtained algebraic characterizations of FO fragments determined by the number of permissible variables. We also generalize Simon's theorem on piecewise testable languages by establishing a decidable algebraic characterization of the Boolean closure of the existential-fragment of FO over countable words. More importantly, we have enriched FO with new infinitary quantifiers and established hierarchical block-product based characterization of the resulting extension $\mathrm{FO}[\infty]$. We also show that $\mathrm{FO}[\infty]$ properties can be expressed simultaneously in FO[cut] as well as WMSO. We do not know if the converse also holds. If true, it will provide a syntactic means to describe the semantic class $\mathrm{FO}[\mathrm{cut}] \cap \mathrm{WMSO}$. We have also shown that these natural logical systems can not have a block-product based characterization using a finite basis.

An interesting future direction is to obtain natural block product decompositions for several sublogics of MSO studied in [10], in particular that of FO[cut] and WMSO. This will complement the equational characterizations presented there and provide the linkages, in the spirit of the fundamental Krohn-Rhodes theorem for finite semigroups, between equational and product based algebraic characterizations over countable words.

References

[1] W. Thomas, Languages, automata, and logic, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Volume 3: Beyond Words, Springer, 1997, pp. 389–455. doi:10.1007/978-3-642-59126-6_7.

[2] J. Pin, Syntactic semigroups, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Volume 1: Word, Language, Grammar, Springer, 1997, pp. 679–746. doi:10.1007/978-3-642-59136-5_10.

[3] P. Tesson, D. Thérien, Logic meets algebra: the case of regular languages, Logical Methods in Computer Science 3 (1) (2007). doi:10.2168/LMCS-3(1:4)2007.

[4] J.R. Büchi, On a Decision Method in Restricted Second Order Arithmetic, in: Proceedings of the 1960 International Congress of Logic, Methodology and Philosophy of Science, Stanford University Press, 1960, pp. 1–12, june.

[5] M. O. Rabin, Decidability of second-order theories and automata on infinite trees, Transactions of the American Mathematical Society 141 (1969) 1–35.

[6] S.Shelah, The monadic theory of order, Ann. of Math. 102 (1975) 379–419.

[7] O. Carton, T. Colcombet, G. Puppis, An algebraic approach to MSO-definability on countable linear orderings, The Journal of Symbolic Logic 83 (3) (2018) 1147–1189. doi:10.1017/jsl.2018.7.

[8] M. P. Schützenberger, On finite monoids having only trivial subgroups, Information and Control 8 (2) (1965) 190–194. doi:10.1016/S0019-9958(65)90108-7.

[9] R. McNaughton, S. A. Papert, Counter-Free Automata (MIT research monograph no. 65), The MIT Press, 1971.

[10] T. Colcombet, A. V. Sreejith, Limited set quantifiers over countable linear orderings, in: M. M. Halldórsson, K. Iwama, N. Kobayashi, B. Speckmann (Eds.), Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part II, Vol. 9135 of Lecture Notes in Computer Science, Springer, 2015, pp. 146–158. `doi:10.1007/978-3-662-47666-6\_12`.

[11] A. Manuel, A. V. Sreejith, Two-variable logic over countable linear orderings, in: P. Faliszewski, A. Muscholl, R. Niedermeier (Eds.), 41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26, 2016 - Kraków, Poland, Vol. 58 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016, pp. 66:1–66:13. `doi:10.4230/LIPIcs.MFCS.2016.66`.

[12] D. M. Gabbay, I. Hodkinson, M. Reynolds, Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 1, Oxford University Press, Oxford, 1994.

[13] A. Bès, O. Carton, Algebraic characterization of FO for scattered linear orderings, in: M. Bezem (Ed.), Computer Science Logic, 25th International Workshop / 20th Annual Conference of the EACSL, CSL 2011, September 12-15, 2011, Bergen, Norway, Proceedings, Vol. 12 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2011, pp. 67–81. `doi:10.4230/LIPIcs.CSL.2011.67`.

[14] K. Krohn, J. Rhodes, Algebraic theory of machines. I. Prime decomposition theorem for finite semigroups and machines, Transactions of the American Mathematical Society 116 (1965) 450–464. `doi:10.2307/1994127`.

[15] H. Straubing, Finite automata, formal logic, and circuit complexity, Birkhaüser Verlag, 1994. `doi:10.1007/978-1-4612-0289-9`.

[16] H. Straubing, D. Thérien, Weakly iterated block products of finite monoids, in: S. Rajsbaum (Ed.), LATIN 2002: Theoretical Informatics, 5th Latin American Symposium, Cancun, Mexico, April 3-6, 2002, Proceedings, Vol. 2286 of Lecture Notes in Computer Science, Springer, 2002, pp. 91–104. `doi:10.1007/3-540-45995-2\_13`.

[17] B. Adsul, S. Sarkar, A. V. Sreejith, First-order logic and its infinitary quantifier extensions over countable words, in: E. Bampis, A. Pagourtzis (Eds.), Fundamentals of Computation Theory - 23rd International Symposium, FCT 2021, Athens, Greece, September 12-15, 2021, Proceedings, Vol. 12867 of Lecture Notes in Computer Science, Springer, 2021, pp. 39–52. doi:10.1007/978-3-030-86593-1_3.

[18] B. Adsul, S. Sarkar, A. V. Sreejith, Block products for algebras over countable words and applications to logic, in: 34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019, IEEE, 2019, pp. 1–13. doi:10.1109/LICS.2019.8785669.

[19] J. G. Rosenstein, Linear orderings, Academic press, 1982.

[20] V. Diekert, P. Gastin, M. Kufleitner, A survey on small fragments of first-order logic over finite words, International Journal of Foundations of Computer Science 19 (3) (2008) 513–548. doi:10.1142/S0129054108005802.

[21] H. Straubing, P. Weil, Varieties, in: J. Pin (Ed.), Handbook of Automata Theory, European Mathematical Society Publishing House, Zürich, Switzerland, 2021, pp. 569–614. doi:10.4171/Automata-1/16.

[22] I. Simon, Piecewise testable events, in: H. Barkhage (Ed.), Automata Theory and Formal Languages, 2nd GI Conference, Kaiserslautern, May 20-23, 1975, Vol. 33 of Lecture Notes in Computer Science, Springer, 1975, pp. 214–222. doi:10.1007/3-540-07407-4_23.

[23] J.-É. Pin, Mathematical foundations of automata theory (December 2020).
URL https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf

[24] V. Diekert, M. Kufleitner, G. Rosenberger, U. Hertrampf, Discrete Algebraic Methods, De Gruyter, Berlin, Boston, 2016. doi:10.1515/9783110413335.