Two-Variable Logic over Countable Linear Orderings

Amaldev Manuel and A. V. Sreejith

Chennai Mathematical Institute (CMI), India
{amal, sreejithav}@cmi.ac.in

Abstract

We study the class of languages of finitely-labelled countable linear orderings definable in two-variable first-order logic. We give a number of characterisations, in particular an algebraic one in terms of circle monoids, using equations. This generalises the corresponding characterisation, namely variety DA, over finite words to the countable case. A corollary is that the membership in this class is decidable: for instance given an MSO formula it is possible to check if there is an equivalent two-variable logic formula over countable linear orderings. In addition, we prove that the satisfiability problems for two-variable logic over arbitrary, countable, and scattered linear orderings are NEXPTIME-complete.

1998 ACM Subject Classification F.4.3 Formal Languages

Keywords and phrases $\circ$-monoids, countable linear orderings, FO^2

Digital Object Identifier 10.4230/LIPIcs.MFCS.2016.67

1 Introduction

Countable linear orderings are linear orderings over countable domains. They are of primary interest in the context of satisfiability of logics due to a result of Shelah [24]: the satisfiability problem of monadic second-order (MSO) logic is undecidable over arbitrary linear orderings, and in particular over the Reals. But by Rabin's theorem [18] the problem remains decidable when considered over countable linear orderings. Thus the class of countable linear orderings sets a natural limit to the decidability of satisfiability problem for MSO over linear orderings. This is in sharp contrast with first-order (FO) logic, that has the corresponding question decidable over arbitrary linear orderings. A second and perhaps more important reason why the class of countable linear orderings are interesting is the logic-algebra connection on its subclasses — MSO definable languages over finite words (*resp.* ω-words) are precisely the class of languages definable by finite monoids (*resp.* ω-semigroups, equivalently Wilke algebras) — extends to countable linear orderings: the result due to Carton-Colcombet-Puppis [4] states that MSO definable languages of countable linear orderings are precisely the class of languages of countable linear orderings recognisable by $\circ$-monoids (recalled in the next section).

The principal import of such a connection is well displayed by the seminal theorem of Schützenberger [21]: over finite words, FO definable languages are precisely the languages recognisable by aperiodic finite monoids, in particular the syntactic monoids of FO definable languages are aperiodic. This immediately yields the decidability of membership in the class of FO definable languages: compute the syntactic monoid of the given language and check if it is aperiodic. Since the time of Schützenberger numerous logics have been characterised algebraically, over finite words, ω-words etc.

However, unlike finite words or ω-words, characterising a logic over countable linear orderings has the following added advantage: An algebraic, in particular decidable, characterisation

41st International Symposium on Mathematical Foundations of Computer Science (MFCS 2016).
Editors: Piotr Faliszewski, Anca Muscholl, and Rolf Niedermeier; Article No. 67; pp. 67:1–67:21
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of a class of languages of countable linear orderings (for instance languages definable by FO) in terms of $\circ$-monoids, immediately provides decidable characterisations over restricted classes of countable linear orderings that are equationally definable (for instance finite words, ω-words, bi-infinite words, rationals etc.). In that sense, characterising a logic algebraically over the class of countable linear orderings in *one shot* characterises it over all equationally definable subclasses.

An elaborate study over a variety of sublogics over countable linear orderings was done in [6] where FO, FO[cut], WMSO, WMSO[cut], MSO[ordinals], MSO[scattered] etc. were characterised algebraically. These characterisations show that WMSO with "cut" quantifiers are equivalent to those with "ordinal" quantifiers, whereas the rest of the logics are expressively different from each other. The study also gives decidability of membership for all these logics.

As a continuation, in this work we consider the class of languages of countable linear orderings that are definable in two-variable first-order logic (FO^2). Two-variable FO is the fragment of FO with at most two variables x, y. While over abritrary structures FO has an undecidable satisfiable problem, FO^2 has a decidable, low complexity satisfiability problem. Yet FO^2 is expressive enough to contain modal logics. This feature of FO^2 has been thoroughly studied and the decidability of satisfiability has been extended to special classes of structures as well as particular vocabularies. FO^2 has been of significant interest over words (and ω-words) as well. Over finite words, FO^2 definable languages have numerous characterisations [26, 25]: they are precisely the class of languages (1) definable in unary LTL [26, 8], (2) recognisable by 2-way partially ordered DFA [22], (3) definable by turtle expressions [27], and (4) whose syntactic monoids are in the variety DA [26] (a finite monoid is in DA if it is aperiodic and all its regular D-classes are subsemigroups) etc. The last characterisation also gives a decision procedure for membership in the class. Not only that FO^2 languages have numerous characterisations, they also have a rich structure inside them [17]— they form an infinite hierarchy under quantifier alternations that is also decidable as shown recently [12].

Though FO^2 is well understood algebraically over finite words, its algebraic characterisation over countable orderings, in particular over infinite ones, is not immediate. This is because even with two variables one can express a variety of "infinitary" conditions: clearly with two variables we can express that letter a has a minimum occurrence (for instance by the formula $\varphi_1 = \exists y \forall x\, (a(x) \wedge a(y) \wedge x \geq y))$, as well as its negation, that is there is an infinite descending chain of a's. Consider the following formula φ_2 that says that if an a-position has an a-position before it, then it has two a-positions before it.

$$\varphi_2 = \forall x\, (a(x) \wedge \exists y\; (a(y) \wedge x > y) \rightarrow \exists y\, (a(y) \wedge x > y \wedge \exists x\, (a(x) \wedge y > x)))$$

The word aa, as well as a^{ω^*} (the ordering $(\mathbb{Z}^-, <)$ labelled with a) does not satisfy $\varphi_1 \wedge \varphi_2$ while the words a and aa^{ω^*} satisfy $\varphi_1 \wedge \varphi_2$. Thus, as $\varphi_1 \wedge \varphi_2$ exemplifies, with two variables one can stipulate both a minimum occurrence as well as existence of a descending chain of a letter. Therefore for the algebraic characterisation of FO^2 one has to make an intricate analysis of whether the letters appear as a minimum or as an infinite chain at different factors of the word.

In the rest of the section, we mention works that are related to the present paper and our contributions.

Related Work

Algebraic characterisations, in particular for FO, for scattered linear orderings are given in [1, 2, 5]. The connection between MSO over countable linear orderings and $\circ$-monoids was proved in [4]. It showed that MSO is equivalent to $\circ$-monoids. This gives an alternate proof of decidability of MSO over countable linear orderings. Moreover it showed that MSO collapses to the second level of the quantifier alternation hierarchy. An algebraic classification of MSO under various forms of set quantifications, in particular corresponding to the sublogics FO, FO[cut], WMSO, WMSO[cut], MSO[ordinals], MSO[scattered], was done in [6].

The literature on FO^2 over arbitrary structures is extensive and we don't mention it here. FO^2 over finite words as well as ω-words has been studied extensively [22, 26, 8, 25, 27, 12, 17]. A survey of various characterisations of FO^2 is given in [25]. The quantifier alternation hierarchy on FO^2 was proved in [11] and the decidability of the hierarchy was shown in [12].

Satisfiability of FO^2 over arbitrary structures were shown to be NEXPTIME-complete in [10]. The corresponding results (also NEXPTIME-complete) was shown for ω-words in [8], and for ordinals in [15]. More recently the satisfiability problem was studied for words with additional linear orderings/preorderings [3, 23, 14, 13].

Satisfiability of LTL over countable linear orderings is PSPACE-complete [7, 19].

Contributions

We study the two variable fragment of first order logic over countable linear orderings and give a number of different characterisations. The simplest characterisation is in terms of temporal logic (TL): FO^2 is equivalent to TL with only the modalities *Future* ($\mathtt{F}$) and *Past* ($\mathtt{P}$). Our major contribution is an algebraic characterisation for FO^2. We show that it corresponds to a subclass of $\circ$-monoids and give two algebraic characterisations for this subclass: (1) by equations, and (2) as the class of $\circ$-monoids that are aperiodic and whose regular $\mathcal{J}$ classes are sub $\circ$-monoids. It follows that the membership in the class is decidable.

Next we study the satisfiability problem for FO^2 over countable linear orderings. The models of FO^2 formulas could be infinite, but we show that a satisfiable formula always admits a scattered model that has a finite representation of small (exponential in the size of the formula) size. Thus we prove that the satisfiability of FO^2 over countable linear orderings is NEXPTIME-complete. From this we also deduce that the satisfiability problems for FO^2 over arbitrary and scattered orderings are NEXPTIME-complete.

Structure of the paper

In Section 2, we introduce words over countable linear orderings, two-variable first-order logic, and the algebra required to characterise FO^2, namely $\circ$-monoids. In Section 3 we prove our main result (Theorem 8) which characterises FO^2. Section 4 deals with the satisfiability of FO^2 over countable linear orderings. Finally we conclude our results in Section 5.

2 Preliminaries

In this section we recall the basic facts about (countable) linear orderings, $\circ$-monoids, logics and related notions.

Words over countable linear orderings. A *linear ordering* $\alpha = (Z, <)$ is a set Z equipped with a total order $<$. For $X, Y \subseteq Z$ we write $X < Y$ if $x < y$ for each x in X and y in Y. In particular $\emptyset < X < \emptyset$ for any set X. Also if $X < Y$, $Y < Z$ and Y is nonempty, then $X < Z$. A *cut* of the linear ordering α is a pair (Z_1, Z_2) such that $Z = Z_1 \cup Z_2$ and $Z_1 < Z_2$. The set of all cuts are linearly ordered and has the least upper bound property [2]. A set L is a *prefix* of X if $X = L \cup K$ and $L < K$ for some $K \subseteq X$. Similarly if $X = L \cup K$ and

$L < K$, then K is a *suffix* of X. Element $z \in Z$ is an *upperbound* (*resp. lowerbound*) of a set $X \subseteq Z$ if $x \leq z$ (*resp.* $z \leq x$) for each x in X. A set X is *right-open* (*resp. left-open*) if it has no maximum element (*resp.* minimum element). Nonempty suffixes of right-open sets are right-open and nonempty prefixes of left-open sets are left-open. The set X is *dense* if between any two elements in the set there is another element; set X is *scattered* if it has no dense subsets. An ordering is a *countable* (scattered) linear ordering if the set Z is countable (scattered). See [20] for further details.

For a finite alphabet A and a linear ordering $\alpha = (Z, <)$, we define a *word* $w : \alpha \to A$ to be a mapping from the set Z to A. We call α the *domain* of w, $dom(w)$. For a word w, we say a point/position x to denote an element $x \in dom(w)$. The notation $w[x]$ denote the letter at the x^{th} position in w. A word has a minimal (respectively maximal) element if its domain has a minimal (maximal) element. The word u is a suffix (prefix) of w if $dom(u)$ is a suffix (prefix) of $dom(w)$. If u and v are words, then uv denotes the unique word w such that $(dom(u), dom(v))$ is a cut of $dom(w)$. This operation is naturally extended to a set of words $\{w_i\}_\alpha$ indexed by a linear ordering α as $\prod_{i\in\alpha} w_i$ (see [6] for more details). For a set $S \subseteq A$, and a word w, we denote the restriction of w to the positions labelled by S as $w_{|S}$. That is $w_{|S} = \{i \in dom(w) \mid w[i] \in S\}$.

The following words are of special interest. ϵ stands for the empty word (the word over an empty domain). The word $\{a\}^\omega$ (denoted in short as a^ω) denotes the word over the domain $(\mathbb{N}, <)$ such that every position is mapped to the letter a. Similarly 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 $S \subseteq A$ of letters, denoted by S^η, is the word over domain $(\mathbb{Q}, <)$ such that any nonempty open interval contains each of the letters in S. This is a unique word (up to isomorphism) (see [4]) and is an example of a dense word, i.e. a word whose domain is dense.

For an alphabet A, the set of all words over nonempty countable domains is denoted by A°. For a word w, we define *alphabet*(w) to be the set of all letters in w. A *language* over the alphabet A is a subset of A°. The language $\{a\}^\infty \subseteq \{a\}^\circ$ (or written as a^∞) denotes all words which are right open. Similarly for a set $S \subseteq A$, the language S^∞ is the set of all words whose letters come only from S and any letter from S can be seen arbitrarily towards the right. The sets $a^{-\infty}$ and $S^{-\infty}$ are defined analogously.

Circle monoids and algebras. A $\circ$-*semigroup* $\mathbf{M} = (M, \pi)$ consists of a set M with an operation $\pi : M^\circ \to M$ which satisfies the following two properties (1) $\pi(a) = a$ for all $a \in M$, (2) *generalised 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 $\mathbf{M}$ has an identity element, then it is called a $\circ$-*monoid.* An element $e \in \mathbf{M}$ is an *idempotent* if $\pi(ee) = e$.

For the rest of the paper, we assume that the monoid $\mathbf{M}$ is finite, that is M is a finite set. The product π is over countable linear orderings and hence it is not possible to finitely represent π. Fortunately, we are able to represent this by a $\circ$-algebra that uses only finite sets and finitely many operations. The following operations are derivable from a $\circ$-monoid $\mathbf{M} = (M, \pi)$:

- *Finite product*, $\cdot : M^2 \to M$ such that $\cdot(a, b) = \pi(ab)$
- *Omega*, $\omega : M \to M$ such that $\omega(a) = \pi(a^\omega)$
- *Omega**, $\omega^* : M \to M$ such that $\omega^*(a) = \pi(a^{\omega^*})$
- *Shuffle*, $\eta : \mathcal{P}(M) \to M$ such that $\{a_1, \ldots, a_k\}^\eta = \pi(\{a_1, \ldots, a_k\}^\eta)$

The resulting structure $(M, \cdot, \omega, \omega^*, \eta)$ is called a $\circ$-*algebra* if it satisfies some additional axioms relating the operations (for example $a \cdot a^\omega = a^\omega$, $(a^\eta)^\omega = a^\eta$ etc.). We skip these details and refer the reader to the paper by Carton et. al [4] for a detailed discussion. The relavant fact is that, for any $\circ$-monoid there exists a unique $\circ$-algebra and vice versa [4].

An important "tool" to understand finite monoids (in our case $\circ$-monoids) is *Green's relations*. In a $\circ$-monoid $\mathbf{M}$, we say that two elements $u \geq_{\mathcal{J}} v$ if there exists two elements $x, y \in \mathbf{M}$ such that $v = xuy$ and $u\mathcal{J}v$ (called J equivalent) if it is both $u \geq_{\mathcal{J}} v$ and $v \geq_{\mathcal{J}} u$. We also say that two elements are $u \geq_{\mathcal{R}} v$ (similarly $u \geq_{\mathcal{L}} v$) if there exists an element $x \in \mathbf{M}$ such that $v = ux$ ($v = xu$). Also $u\mathcal{R}v$ if $u \geq_{\mathcal{R}} v$ and $v \geq_{\mathcal{R}} u$. Similarly we can define $u\mathcal{L}v$. The relations $\mathcal{L}$ and $\mathcal{R}$ are right and left congruences respectively. If a $\mathcal{J}$ class contains an idempotent then it is called a *regular* $\mathcal{J}$ class. All elements in a $\mathcal{J}$ class can be described by an "eggbox" structure, such that $u\mathcal{J}v$ iff there exists elements $x, y \in \mathbf{M}$ such that $u\mathcal{R}x\mathcal{L}y\mathcal{R}v$. For a more detailed elaboration on this subject see [16].

The class of $\circ$-monoids that satisfies the property — there exists an $n \in \mathbb{N}$ such that $a^n = a^{n+1}$ for all $a \in \mathbf{M}$ — are called *aperiodic*. It is precisely the class of $\circ$-monoids which do not contain any non-trivial group as a subsemigroup of $(M, \cdot)$ (by Schützenberger's theorem [21]).

One way to denote a class of $\circ$-monoids is by equations. For instance, we say that $\mathbf{M}$ satisfies the equation $x^* = x^{\omega} x^{\omega^*}$, if for all elements $a \in \mathbf{M}$, $a^* = a^{\omega} a^{\omega^*}$, where a^* is the unique idempotent power of a.

We say that a language $L \subseteq A^{\circ}$ is recognised by the $\circ$-monoid $\mathbf{M}$, if there is a morphism, $\gamma : A^{\circ} \to \mathbf{M}$ and a subset $S \subseteq \mathbf{M}$ such that $L = \gamma^{-1}(S)$. The *syntactic $\circ$-monoid* of a language L is the minimal $\circ$-monoid $\mathbf{M}$ recognising L that has the following universal property: any $\circ$-monoid recognising L has a morphism onto $\mathbf{M}$.

Logics. Monadic second-order logic (MSO) over a finite alphabet A is a logic which can be inductively built using the following operations.

$$a(x) \mid x < y \mid x = y \mid \alpha_1 \vee \alpha_2 \mid \neg\alpha \mid x \in X \mid \exists x\ \alpha \mid \exists X\ \alpha$$

Here $a \in A$. If we remove the second-order quantification, we get first-order logic (FO). If we further restrict the logic to use only two variables (but allowing repetitions) we get FO^2. Note that, we do not have the *successor* relation in our logic.

A formula with no free variables is called a sentence. The language of a sentence φ (denoted by $L(\varphi)$) is the set of all $u \in A^{\circ}$ that satisfies φ.

Over finite words, FO^2 can talk about occurrence of letters and also about the order in which they appear [8, 27]. Over countable linear orders, FO^2 can also talk about an infinite sequence of a letter. For example, the language a^{∞} is definable in FO^2 by stating that, every position is labelled by a and there is no maximum position.

$$\big(\forall x\ \exists y > x\big) \wedge \big(\forall x\ a(x)\big)$$

Also, for a subset $S \subseteq A$, we can also express the language S^{∞} in FO^2.

$$\big(\forall x \bigwedge_{a \in S} \exists y > x\ a(y)\big) \wedge \big(\forall x \bigvee_{a \in S} a(x)\big)$$

Analogously, FO^2 can also talk about left open words.

The temporal logic $\{\mathtt{F}, \mathtt{P}\}$-TL over the alphabet A is the logic with the set of formulas — a when a is a letter in A, and $\mathtt{F}\varphi$ and $\mathtt{P}\varphi$ when φ is a formula — that is closed under Boolean operations. To state the semantics fix a word $u \in A^{\circ}$. A position $i \in \mathit{dom}(u)$ satisfies — the formula a if i is labelled with the letter a, and the formula $\mathtt{F}\varphi$ (*resp.* $\mathtt{P}\varphi$) if there is a position $i < j \in \mathit{dom}(u)$ (*resp.* $i > j \in \mathit{dom}(u)$) that satisfies the formula φ. The semantics for Boolean connectives are defined in the usual way. The word u satisfies the formula φ if there is a position $i \in dom(u)$ that satisfies the formula (see [8] for a detailed presentation). The language of the formula φ is the set of all $u \in A^\circ$ that satisfies φ.

3 Characterisation

In this section, we give the algebraic characterisation for FO^2(<) over countable linear orderings. As we noted earlier, ∘-monoid captures MSO. Here we identify a subclass which will capture the two-variable first-order fragment. Our characterisation builds on the characterisation for FO^2 on finite words given in [26]. In particular, we crucially use a generalisation of the congruence given there.

▶ **Definition 1.** We define ∘-DA to be the subclass of ∘-monoids that satisfy the following equations.

1. $(xyz)^*y(xyz)^* = (xyz)^*$

2. $x^* = (x)^\omega (x)^{\omega^*}$

3. $\{x_1, \ldots, x_k\}^\eta = (x_1 \cdots x_k)^{\omega^*} (x_1 \cdots x_k)^\omega$

The first equation corresponds to the variety DA of finite monoids [25]. It identifies the constraints the product operation has to satisfy. The second equation corresponds to FO definable languages of countable linear orderings [6]. This equation states that a $\mathcal{J}$ class with an idempotent will also contain its omega and omega* powers. The last equation says that, ∘-DA cannot differentiate between dense and scattered orderings.

The connection between logic and algebra is established using the following congruence.

A congruence on words

Let $u \in A^\circ$ be an arbitrary word. $alphabet(u)$ is defined as the set of all letters occurring in u. For a letter a in $alphabet(u)$, let $P_u(a)$ denote the set of all positions in u labelled with a. Let $T_r^1(u) \subseteq alphabet(u)$ be the set of all letters a such that $P_u(a)$ has a maximal element. Furthermore, let $T_r^\omega(u)$ be the set $alphabet(u) \setminus T_r^1(u)$, i.e. the set of all letters that do not have a maximal occurrence. Similarly let $T_l^1(u) \subseteq alphabet(u)$ be the set of all letters a such that $P(a)$ has a minimal element, and let $T_l^{\omega^*}(u)$ be the set $alphabet(u) \setminus T_l^1(u)$.

▶ **Definition 2.** The relation $\lesssim_r$ over the set of letters $T_r^\omega(u)$ is defined as follows:

$a \lesssim_r b$ if each a-position i in u has a b-position j to its right (i.e. $j > i$).

▶ **Lemma 3.** *The relation $\lesssim_r$ is a total preorder on the set $T_r^\omega(u)$.*

Proof. Since for each letter a in $T_r^\omega(u)$ the set $P_u(a)$ does not have a maximum, clearly $a \lesssim_r a$. Next, let $a, b, c \in T_r^\omega$ be such that $a \lesssim_r b$ and $b \lesssim_r c$. By definition of $\lesssim_r$, every a-position has a b-position to the right, which in turn has a c-position to its right. Hence $a \lesssim_r c$. It only remains to show that $\lesssim_r$ is total. Assume $a \not\lesssim_r b$. By definition, there is an a-position i that has no b-position to its right. Hence all b-positions lie to the left of i, and therefore every b-position has an a to its right, and therefore $b \lesssim_r a$. ◀

We write $\sim_r$ to denote the equivalence relation associated with the preorder $\lesssim_r$. For a letter a in $T_r^\omega(u)$ we let $[a]_r \subseteq T_r^\omega(u)$ denote the equivalence class of a with respect to the total preorder $\lesssim_r$, i.e. $[a]_r = \{b \in T_r^\omega(u) : b \sim_r a\}$. Also, we extend the definition of P_u to equivalence classes by defining $P_u([a]_r) = \bigcup_{a \in [a]_r} P_u(a)$. We write $<_r$ to denote the total order on $\{[a]_r : a \in T_r^\omega(u)\}$.

By symmetry, the dual relation $\lesssim_l$ defined as,

$b \lesssim_l a$ if each a-position in u has a b-position to its left,

is also a total preorder. The corresponding equivalence relation and strict order relation are denoted as $\sim_l$ and $<_l$. Given a circle word u the preorders $\lesssim_r$ and $\lesssim_l$ associated with u are called the *right preorder* and *left preorder* of u respectively. As before we define $P_u([a]_l) = \bigcup_{a \in [a]_l} P_u(a)$.

▶ **Example 4.** Let $S = \{a, b\}$ and let $u \in S^{-\infty}$ be an arbitrary word. Consider the word $v = ua^{\omega^*}a^{\omega}ab^{\omega} \in \{a,b\}^{\circ}$. Then $T_l^1(u) = T_l^1(v) = \emptyset$ and $T_l^{\omega^*}(u) = T_l^{\omega^*}(v) = \{a, b\}$, since a and b occur infinitely often towards left in both u and v. It also follows that $a \lesssim_l b$ and $b \lesssim_l a$. Since u is an arbitrary word, we do not know about $T_r^1(u)$ and $T_r^{\omega}(u)$. But, since a has a maximum point in v, we have $T_r^1(v) = \{a\}$ and $T_r^{\omega}(v) = \{b\}$. Moreover $b \lesssim_r b$.

Consider another word $w = a^{\omega}b^{\omega}$. Here we have $T_l^1(w) = \{a, b\}$ and $T_l^{\omega^*}(w) = T_r^1(w) = \emptyset$. We also have $T_r^{\omega}(w) = \{a, b\}$ and $a \lesssim_r b$ but $b \not\lesssim_r a$.

We will now introduce left/right decomposition of words. The idea is to factorise a word in a particular way to capture the "pivot" points for an FO^2 formula.

▶ **Definition 5.** Let $a \in \mathit{alphabet}(u)$. If $a \in T_l^1(u)$, then there exists a unique factorisation of u as (u_0, a, u_1) such that $u = u_0 a u_1$ and $a \notin \mathit{alphabet}(u_0)$. This is called the a-left decomposition of u. Similarly there is a unique factorisation of u as (u_0, a, u_1) such that $a \notin \mathit{alphabet}(u_1)$, if $a \in T_r^1(u)$. This is called the a-right decomposition of u.

We are also interested in left decomposition obtained by a set of positions $P_u([a]_l)$, where $[a]_l \in T_l^{\omega^*}(u)/\sim_l$. That is for a subset of positions $P_u([a]_l)$ of u, we define the $P_u([a]_l)$*-left decomposition of a word* u to be the unique maximal cut (u_0, u_1) such that $P_u([a]_l) \cap \mathit{dom}(u_0) = \emptyset$. Note that if $S = \{b \mid b \sim_l a\}$, then there is a prefix of u_1 such that $u_1 \in S^{-\infty}$. This follows from the fact that, the decomposition (u_0, u_1) is a maximal cut. Similarly the $P_u([a]_r)$*-right decomposition of a word* u is defined to be the unique minimal cut (u_0, u_1) such that $P_u([a]_r) \cap \mathit{dom}(u_1) = \emptyset$.

With the left/right decomposition defined, we can define the *congruence on words*, $\equiv_n$ which essentially captures a sequence of unique decompositions.

▶ **Definition 6.** For an alphabet A, a natural number $n \in \mathbb{N}$ and words $u, v \in A^{\circ}$, we define $u \equiv_n v$ by induction on $m = n + |A|$ as follows.

1. If $n = 0$ (the base case): $u \equiv_0 v$ for all $u, v \in A^{\circ}$.
2. If $n > 0$: We say $u \equiv_n v$ if the following conditions are satisfied:
   - **a.** $\mathit{alphabet}(u) = \mathit{alphabet}(v)$, $T_r^1(u) = T_r^1(v)$, and $T_l^1(u) = T_l^1(v)$. (This condition implies that $T_r^{\omega}(u) = T_r^{\omega}(v)$ and $T_l^{\omega^*}(u) = T_l^{\omega^*}(v)$).
   - **b.** The right preorders of u and v (both on the same set by the previous observation) are the same. Similarly the left preorders of u and v are the same. (We denote the left and right preorders as $\lesssim_l, \lesssim_r$ respectively).
   - **c.** For each $a \in T_l^1(u) = T_l^1(v)$, let (u_0, a, u_1) be the a-left decomposition of u, and let (v_0, a, v_1) be the a-left decomposition of v, then $u_0 \equiv_n v_0$ and $u_1 \equiv_{n-1} v_1$. Note that the induction parameter has reduced in both cases: u_0 has at least one letter less than u; and we have a lesser congruence in u_1.
   - **d.** Similarly, for each $a \in T_r^1(u) = T_r^1(v)$, let (u_0, a, u_1) be the a-right decomposition of u and let (v_0, a, v_1) be the a-right decomposition of v, then $u_0 \equiv_{n-1} v_0$ and $u_1 \equiv_n v_1$.

e. For each class $[a]_l \in T_l^{\omega^*}(u)/\sim_l = T_l^{\omega^*}(v)/\sim_l$, let (u_0, u_1) be the $P_u([a]_l)$-left decomposition of u and let (v_0, v_1) be the $P_v([a]_l)$-left decomposition of v, then $u_0 \equiv_n v_0$ and $u_1 \equiv_{n-1} v_1$. Again, the induction parameter has reduced in both cases: u_0 has at least one letter less than u; and we have a lesser congruence in u_1.

f. Similarly for each class $[a]_r \in T_r^{\omega}(u)/\sim_r = T_r^{\omega}(v)/\sim_r$, let (u_0, u_1) be the $P_u([a]_r)$-right decomposition of u and let (v_0, v_1) be the $P_v([a]_r)$-right decomposition of v, then $u_0 \equiv_{n-1} v_0$ and $u_1 \equiv_n v_1$.

▶ **Lemma 7.** *The relation $\equiv_n$ is a congruence relation for every $n \in \mathbb{N}$.*

Proof. To prove the claim we need to show that for each words $u, v, w \in A^\circ$, if $u \equiv_n v$, then $uw \equiv_n vw$ and $wu \equiv_n wv$. The cases are symmetric and we consider only the first case. We prove the claim using induction on $n + \min(|alphabet(u)|, |alphabet(v)|)$. When $n = 0$, there is nothing to show.

Next assume that $n > 0$ and $u \equiv_n v$. If $\min(|alphabet(u)|, |alphabet(v)|) = 0$ then $u = v = \epsilon$ and the claim is obvious. Otherwise $u \neq \epsilon \neq v$. Since $u \equiv_n v$, it is clear that $alphabet(uw) = alphabet(u) \cup alphabet(w) = alphabet(v) \cup alphabet(w) = alphabet(vw)$. Similarly $T_l^1(uw) = T_l^1(u) \cup \left(T_l^1(w) \setminus alphabet(u)\right) = T_l^1(v) \cup \left(T_l^1(w) \setminus alphabet(v)\right) = T_l^1(vw)$, and $T_r^1(uw) = T_r^1(w) \cup \left(T_r^1(u) \setminus alphabet(w)\right) = T_r^1(w) \cup \left(T_r^1(v) \setminus alphabet(w)\right) = T_r^1(vw)$.

Next we verify that the left-preorders of uw and vw are the same. Assume $b \lesssim_l a$ is in the left preorder of uw. There are two cases to consider, depending on whether $a \notin T_l^{\omega^*}(u)$ or not.

1. $a \notin T_l^{\omega^*}(u)$: Hence $a \in T_l^{\omega^*}(w)$. If $b \in T_l^{\omega^*}(u)$, then by assumption $b \in T_l^{\omega^*}(v)$ and therefore $b \lesssim_l a$ is in the left preorder of vw. If $b \notin T_l^{\omega^*}(u)$, then $b \lesssim_l a$ is in the left preorder of w and therefore $b \lesssim_l a$ is in the left preorder of vw.
2. $a \in T_l^{\omega^*}(u)$: Therefore $b \lesssim_l a$ is in the left preorder of u and therefore of v and vw.

Similarly if $b \lesssim_r a$ is in the right preorder of uw we have that $b \lesssim_r a$ is in the right preorder of vw.

We need to now show that a left/right decomposition is in uw iff it is in vw. We will show if uw has a particular decomposition then vw will have that decomposition such that the congruence on the factors satisfy appropriately. Symmetrically we can show the same claim for vw. This will give us that $uw \equiv_n vw$.

1. Let (u_0, a, w_1) be a a-left decomposition of uw: If $a \in alphabet(u)$, then there is an a-left decomposition of $u = (u_0, a, u_1)$ such that $w_1 = u_1 w$. Therefore, there exists an a-left decomposition of $v = (v_0, a, v_1)$ and hence there is an a-left decomposition of $vw = (v_0, a, v_1 w)$. Since $u \equiv_n v$, we have $u_0 \equiv_n v_0$ and $u_1 \equiv_{n-1} v_1$. By inductive hypothesis, we therefore have $u_1 w \equiv_{n-1} v_1 w$. On the other hand if $a \notin alphabet(u)$, then there is an a-left decomposition of $w = (w_0, a, w_1)$. If follows that $a \notin alphabet(v)$ and therefore there is an a-left decomposition of $vw = (vw_0, a, w_1)$. Its clear that $uw_0 \equiv_n vw_0$.
2. Let (u_0, w_1) be a $P_{uw}([a]_l)$-left decomposition of uw. We will look at two cases
   - **a.** Let $\{a, b\} \subseteq T_l^{\omega^*}(u)$: Then there exists a $P_u([a]_l)$-left decomposition (u_0, u_1) of u such that $w_1 = u_1 w$. Therefore by our assumption, there exists a $P_v([a]_l)$-left decomposition of $v = (v_0, v_1)$ and hence there exists a $P_{vw}([a]_l)$-left decomposition of $vw = (v_0, v_1 w)$ such that $u_0 \equiv_n v_0$ and $u_1 w \equiv_{n-1} v_1 w$.
   - **b.** Let $\{a, b\} \cap T_l^{\omega^*}(u) = \emptyset$: Then $\{a, b\} \cap T_l^{\omega^*}(v) = \emptyset$ and there exists a $P_w([a]_l)$-left decomposition of $w = (w_0, w_1)$. Therefore $vw = (vw_0, w_1)$ is a $P_{vw}([a]_l)$-left decomposition of vw. Its clear that $uw_0 \equiv_n vw_0$.
3. A similar cases analysis can be given for both right decompositions.

We have now showed that $uw \equiv_n vw$. This concludes our proof. ◀

Main theorem

We are now in a position to state our main theorem.

▶ **Theorem 8.** *Let $L \subseteq A^\circ$. Then the following are equivalent:*

1. *L is definable in $\{\mathtt{F},\mathtt{P}\}$-TL.*
2. *L is $\mathrm{FO}^2(A, \leq)$ definable.*
3. *L is a union of $\equiv_n$ congruent classes for some $n \in \mathbb{N}$.*
4. *L is recognised by a $\circ$-DA.*
5. *L is recognised by an aperiodic $\circ$-monoid where all regular $\mathcal{J}$ classes are sub $\circ$-monoids.*
6. *The syntactic $\circ$-monoid of L is in $\circ$-DA.*

The proof of $(1 \Leftrightarrow 2)$ follows easily (see [8, 7]).
In subsection 3.1 we show the equivalence of the different monoid views $(4 \Leftrightarrow 5 \Leftrightarrow 6)$.
In subsection 3.3 we show $(4 \Rightarrow 3)$.
To prove $(2 \Rightarrow 4)$, we use 2-pebble *Ehrenfeucht-Fraïssé* (EF) games [26]. The EF game gives a game congruence $\cong_n$ defined as: $u \cong_n v$ if the duplicator wins the n-round 2-pebble game on the pair of words (u, v). See [26] for the game congruence and its equivalence to FO^2. Thus it suffices to show that the game congruence satisfies the equations of $\circ$-DA.
To show direction $(3 \Rightarrow 2)$ we follow the proof in [26]. It suffices to show that if $L \subseteq A^\circ$ is a union of $\equiv_n$ congruent classes for some n, then it is definable in $\mathrm{FO}^2(<)$. More precisely we prove the following lemma (again using the equivalence of game congruence $\cong_n$ and FO^2).

▶ **Lemma 9.** *For words $u, v \in A^\circ$, If $u \not\equiv_n v$, then $u \ncong_{n+alphabet(u)} v$ i.e. the spoiler has a winning strategy in the 2-pebble $n + alphabet(u)$-round EF game on u and v.*

Since the syntactic $\circ$-monoid (and its finite representation using $\circ$-algebra) is computable given an MSO formula [4], it follows that it is decidable to check whether the language is FO^2 definable.

▶ **Corollary 10.** *For a sentence ϕ in MSO$[<]$, it is decidable whether $L(\phi)$ is FO$^2[<]$ definable.*

In the next subsection we show the equivalence of the different monoid views. The subsection after that shows that if a language is accepted by a $\circ$-monoid, then it is a union of congruence classes $\equiv_n$ for some $n \in \mathbb{N}$.

3.1 The different Monoid views

In this subsection we show that the different views of $\circ$-DA are equivalent. That is, $(4 \Leftrightarrow 5 \Leftrightarrow 6)$ of Theorem 8. The direction $(4 \Rightarrow 5)$, follows from standard ideas in semigroup theory

Proof of Theorem 8, $(4 \Rightarrow 5)$. Let $\mathbf{M}$ be a $\circ$-monoid, which satisfies the (3) equations in Definition 1. First we show that $\mathbf{M}$ is aperiodic. This follows from, equation (2), since $a^n = a^\omega a^{\omega^*} = a(a^\omega a^{\omega^*}) = a^{n+1}$.

We need to now show that all regular $\mathcal{J}$ classes of $\mathbf{M}$ are sub $\circ$-monoids. In other words, we need to show that a regular $\mathcal{J}$ class is closed under operations, finite product, omega, omega* and shuffle. Let J be a regular $\mathcal{J}$ class. From equation (1) of $\circ$-DA it follows that J is closed under finite product. Let $x \in J$. We will show that $x^\omega \mathcal{R} x$. Since J is closed under product, $x^n \mathcal{R} x$. From equation (2), $(x^n)\mathcal{R}(x^n)^\omega$. Similarly we that J is closed under omega*. We need to now show that J is closed under shuffle operation. Consider the element $t = \{a_1, \ldots, a_k\}^\eta$ where $a_1, \ldots, a_k \in J$. By equation (3), we get that

$t = (a_1 \ldots a_k)^{\omega^*}(a_1 \ldots a_k)^{\omega}$. Since J is closed under concatenation/omega/omega* operation, it follows that $t \in J$ and hence J is closed under shuffle too. ◀

The reverse direction ($5 \Rightarrow 4$), follows from the below lemma:

▶ **Lemma 11.** *Let $\mathbf{M}$ be an aperiodic $\circ$-monoid such that all regular $\mathcal{J}$ classes of $\mathbf{M}$ are sub $\circ$-monoids. Let $\gamma : A^\circ \to \mathbf{M}$ be a morphism and $u \in A^\circ$, such that $\gamma(u) = e$ an idempotent. Then, for all words $v \in \{alphabet(u)\}^\circ$, we have $\gamma(uvu) = \gamma(u)$.*

Proof. Consider the evaluation tree of the word v. The evaluation tree is a bounded height tree. Each node in the tree represents the morphism of a factor of v. The value at the root of the tree is $\gamma(v)$. For a detailed study on evaluation trees, refer [4].

We will now inductively show the following property in the tree. If t is the value at a particular node, then $ete = e$. We are done, once we show this, since the value at the root node is $\gamma(v)$ and therefore $e\gamma(v)e = e$. Our proof depends on the type of the node.

- Leaf node: Let the node be the letter $a \in alphabet(u)$. Without loss of generality let us denote by $\gamma(a) = a$. Therefore $e = xay$. Hence $e = ee \;\mathcal{J}\; exa \;\mathcal{J}\; exae \;\mathcal{J}\; ae \;\mathcal{J}\; eae$. Here we used the fact that the $\mathcal{J}$ class is regular (since idempotent e is in that class) and therefore closed under finite product. Since $e \geq_{\mathcal{R}} eae$ and $e \geq_{\mathcal{L}} eae$, we have $e = eae$.
- Value of node $t = l.r$: From IH, $el\mathcal{J}re\mathcal{J}e$. Hence $elre = e$.
- Value of node $t = f^\omega$: By IH, $efe = e$ and $f\mathcal{J}f^\omega$. Therefore $f^\omega x = f$, for an $x \in \mathbf{M}$ and hence $e = efe\mathcal{J}ef^\omega\mathcal{J}ef^\omega e$. Therefore $ete = e$.
- Value of node $t = f^{\omega^*}$: By IH, $efe = e$ and $f\mathcal{J}f^{\omega^*}$. Therefore $xf^{\omega^*} = f$, for an $x \in \mathbf{M}$ and hence $e = efe\mathcal{J}f^{\omega^*}e\mathcal{J}ef^{\omega^*}e$. Therefore $ete = e$.
- Let value of node $t = \{a_1, \ldots, a_k\}^\eta$: Let $S = \{a_1, \ldots a_k\}$ and $f = (a_1 \ldots a_k)^n$. Our aim is to show that $S^\eta = f^{\omega^*}f^\omega$. Since S^η and f are idempotents they are members of a regular $\mathcal{J}$ class. The following relations follow from equations of $\circ$-monoids (one can also view it as from uniqueness of perfect shuffle) [4]. $f\mathcal{J}f^\eta\mathcal{J}a_1f^\eta\mathcal{J}\{a_1, f\}^\eta\mathcal{J}\{S \cup f\}^\eta\mathcal{J}\{f, S^\eta\}^\eta\mathcal{J}(fS^\eta)^\eta\mathcal{J}fS^\eta\mathcal{J}S^\eta$.
  Since $f\mathcal{J}S^\eta$, we have, $t = S^\eta = f^{\omega^*}f^\omega$. Therefore $eS^\eta e = ef^{\omega^*}f^\omega e$. The claim now follows from the previous 4 cases.

We have shown that for all t in some node $ete = e$. This proves our claim. ◀

Using the above lemma, we can show the following direction of Theorem 8.

Proof of Theorem 8, ($5 \Rightarrow 4$). Let $\mathbf{M}$ be an aperiodic $\circ$-monoid such that all regular $\mathcal{J}$ classes are sub $\circ$-monoids. We show that all equations of $\circ$-DA given in Definition 1 are satisfied.

1. Equation $(xay)^n a (xay)^n = (xay)^n$. Since $(xay)^n = e$ is an idempotent, the equation follows from Lemma 11.
2. Equation $x^n = (x^n)^\omega (x^n)^{\omega^*}$: Since $x^n = e$ is an idempotent, follows from Lemma 11.
3. Equation $\{x_1, \ldots, x_k\}^\eta = (x_1 \ldots x_k)^{\omega^*}(x_1 \ldots x_k)^\omega$: Since both elements are in a regular $\mathcal{J}$ class, the equation follows from Lemma 11.

◀

To prove direction ($4 \Rightarrow 6$), assume L is recognised by a monoid in $\circ$-DA. Since, $\circ$-DA is closed under quotienting, it follows that the syntactic monoid of L satisfies the equations of $\circ$-DA (see [6] for more details about syntactic congruence and monoids).

3.2 Congruence on words to Logic

▶ **Lemma 9.** *For words $u, v \in A^\circ$, If $u \not\equiv_n v$, then $u \ncong_{n+alphabet(u)} v$ i.e. the spoiler has a winning strategy in the 2-pebble $n + alphabet(u)$-round EF game on u and v.*

Proof. We prove the claim by induction on $n + alphabet(u)$. When $n = 0$, then all words $u, v \in A^\circ$ are $\equiv_0$-equivalent, and the claim is vacuously true. Now assume $n > 0$ and $u, v \in A^\circ$ are such that $u \not\equiv_n v$. We have several cases.

If $alphabet(u) \neq alphabet(v)$ then Spoiler wins the game in 1 move: Without loss of generality, she picks a letter $a \in alphabet(u) \setminus alphabet(v) \neq \emptyset$ and Duplicator has no successful response since the word v does not contain the letter a.

If $T_l^1(u) \neq T_l^1(v)$, then spoiler wins the game in 2 moves: Without loss of generality assume that $T_l^1(u) \setminus T_l^1(v) \neq \emptyset$ and let $a \in T_l^1(u) \setminus T_l^1(v)$. Spoiler picks the minimum occurrence of the letter a in u. Duplicator picks a letter a in v (otherwise she immediately loses) and Spoiler responds by picking an a-position to its left. Duplicator does not have an a-position to pick in the word u to the left of the pebble already in u and she loses. The case of $T_r^1(u) \neq T_r^1(v)$ is similar.

Next assume that $alphabet(u) = alphabet(v)$ and $T_l^1(u) = T_l^1(v)$ (hence $T_l^{\omega^*}(u) = T_l^{\omega^*}(v)$), but the left-preorders of u and v differ. Then there is a pair $b \lesssim_l a$ that is in one of the preorders, but not in the other, for some $a, b \in T_l^{\omega^*}(u) = T_l^{\omega^*}(v)$. Without loss of generality assume that $b \lesssim_l a$ is in the left-preorder of u, but not in the left-preorder of v. Therefore, in u each a-position has a b-position to its left, but not in v. Hence in v, there exists an a-position that is to the left of all the b-positions. We claim that the Spoiler has a winning strategy in 2 moves: She picks an a-position in v that has no b to the left, the Duplicator has to respond by picking an a-position in u. The Spoiler then pebbles a b-position to the left of the a-position picked by the Duplicator. Duplicator has no successful move and she loses. The case when $alphabet(u) = alphabet(v)$ and $T_r^1(u) = T_r^1(v)$, but the right-preorders of u and v differ, is similar.

Next assume that $a \in T_l^1(u) = T_l^1(v)$, and $(u_0, a, u_1), (v_0, a, v_1)$ are the a-left decompositions of u and v respectively, but either $u_0 \not\equiv_n v_0$ or $u_1 \not\equiv_{n-1} v_1$. Let $dom(u_0) < \{i\} < dom(u_1)$ and $dom(v_0) < \{i'\} < dom(v_1)$ be the positions corresponding to the a's sandwiched between u_0, u_1 and v_0, v_1 respectively. We have two cases. For the first case, assume that $u_0 \not\equiv_n v_0$. Then by induction hypothesis Spoiler has a winning strategy in the $n + \alpha(u) - 1$ round game on (u_0, v_0). For the game on u_0 and v_0, Spoiler mimics this strategy and wins, or at some point Duplicator places a pebble in a position that is not in u_0 or v_0. Let $j \in u_1$ be the position pebbled by the Duplicator; j cannot be i since v_0 does not contain any a-position (and hence will be immediately losing). In the following move, Spoiler pebbles the position i in u and Duplicator is forced to respond with an a-position in v_0 and loses. For the second case, assume that $u_1 \not\equiv_{n-1} v_1$. Then by induction hypothesis Spoiler has a winning strategy in the $n - 1 + \alpha(u)$ round game on (u_1, v_1). Again, Spoiler follows this strategy until she wins the game or at some point Duplicator pebbles a position (say using the red pebble) in $(dom\{u_0\} \cup \{i\}) \bigcup (dom\{v_0\} \cup \{i'\})$. Without loss of generality assume the position is in $dom\{u_0\} \cup \{i\}$, then Spoiler responds by pebbling the position i' using the blue pebble. Duplicator is forced to respond by pebbling an a-position in u_0 and loses.

The case for a-right decompositions follows from symmetry.

Next assume that $T_l^{\omega^*}(u) = T_l^{\omega^*}(v)$ and the left-preorders of u and v are the same. Furthermore, assume that $[a]_l \in T_l^{\omega^*}(u)/\sim_l = T_l^{\omega^*}(v)/\sim_l$, and let (u_0, u_1) be the $P_u([a]_l)$-left decomposition of u and let (v_0, v_1) be the $P_v([a]_l)$-left decomposition of v, but either $u_0 \not\equiv_n v_0$ or $u_1 \not\equiv_{n-1} v_1$. We have two cases. For the first case, assume that $u_0 \not\equiv_n v_0$. Then by induction hypothesis Spoiler has a winning strategy in the $n+alphabet(u)-1$ game on u_0 and v_0. Spoiler follows this winning strategy in the game on u_0 and v_0, either winning the game or at some point Duplicator puts a pebble in u_1 or v_1 (say on u_1 without loss of generality). Then Spoiler responds by placing the other pebble on an a-position to its left, where $a \in [a]_l$. Such a position is guaranteed to exist by definition of $T_l^{\omega^*}(u)$ and the cut (u_0, u_1). Duplicator is forced to respond by pebbling an a-position in v_0, and loses the game since v_0 does not contain any a-position. For the second case, assume that $u_1 \not\equiv_{n-1} v_1$. Again, by induction hypothesis Spoiler has a winning strategy in the $n-1+alphabet(u)$-round game on u_1 and v_1. Spoiler plays according to this winning strategy, either winning the game, or Duplicator places a pebble (say blue pebble) at a position in u_0 or v_0 at some moment (say in u_0 without loss of generality). Then spoiler responds by placing the red pebble on an a-position in v_1 that is to the left of the blue pebble in v_1, where $a \in [a]_l$. Duplicator cannot replicate this move in u_0 since all a-position are in u_1 and she loses.

The case of right-preorder classes is similar. This concludes the all the cases and therefore $u \not\equiv_n v$. ◀

3.3 Algebra to Congruence

In this subsection we show direction (4 ⇒ 3) of Theorem 8. The proof improves on the equivalence of the congruence and algebra given in [26]. We show that a language recognisable by a ∘-monoid in ∘-DA, satisfies the congruence relation $\equiv_n$ for some $n \in \mathbb{N}$. Let L be recognised by the morphism $\gamma : A^\circ \to \mathbf{M}$, where $\mathbf{M}$ is in ∘-DA. It suffices to show that there exists an $n \in \mathbb{N}$ such that $\equiv_n$ is a finer congruence than the monoid congruence. That is for $u, v \in A^\circ$, if $u \equiv_n v$, then $\gamma(u) = \gamma(v)$. Since $\mathbf{M}$ is an aperiodic monoid (follows from equations of ∘-DA) it is sufficient to show that $u\mathcal{R}v$ and $u\mathcal{L}v$.

The left/right decomposition of words are closely related to how the $\mathcal{R}$ classes fall in the word. The following definition identifies a sequence of $\mathcal{R}$-smooth factors (those factors where there is no $\mathcal{R}$ fall), and the subsequent lemma shows there exists such a unique sequence.

▶ **Definition 12.** Let $\gamma : A^\circ \to \mathbf{M}$. Let $w \in A^\circ$. Then the $\mathcal{R}$ decomposition of w is defined as the sequence $(w_0, a_1, w_1, a_2, \ldots, a_k, w_k)$ such that

1. $a_i \in A \cup \{\epsilon\}$ and $w_i \in A^*$, for all $i \leq k$.
2. $w = w_0 a_1 \ldots a_k w_k$.
3. For each $0 < i \leq k$, if a_i is empty, then the following conditions hold:
   a. w_i does not have a left end point.
   b. $(w_0 a_1 \ldots a_i w_i') \mathrel{\mathcal{R}} \gamma(w_0 a_1 \ldots a_i w_i)$, for all nonempty prefix w_i' of w_i.
   c. $\gamma(w_0 a_1 \ldots w_{i-1}) \not\mathrel{\mathcal{R}} \gamma(w_0 a_1 \ldots w_{i-1} a_i w_i)$.
4. For each $0 < i \leq k$, if a_i is not empty, then the following holds:
   a. $\gamma(w_0 a_1 \ldots a_i) \mathrel{\mathcal{R}} \gamma(w_0 a_1 \ldots a_i w_i)$.
   b. $\gamma(w_0 a_1 \ldots w_{i-1}) \not\mathrel{\mathcal{R}} \gamma(w_0 a_1 \ldots w_{i-1} a_i)$.

▶ **Lemma 13.** *Let $w \in A^\circ$ be an arbitrary word. Then, there is a unique $\mathcal{R}$ decomposition $(w_0, a_1, \ldots, a_k, w_k)$ of w.*

Proof. Let $w \in A^\circ$. By a sequence of operations we get the factors $w_0, a_1, \ldots, w_k$. We first get w_0 and a_1 and then show how to inductively built other w_i's and a_i's.

Let (w_0, w_0') be a factorization of w such that w_0 is the maximal prefix of w where w_0 is $\mathcal{R}$ smooth. If w_0' is not left-open, let $w_0' = a_1 w_0''$. Otherwise take $a_1 = \epsilon$. Rewrite w_0' with w_0'' if $a_i \neq \epsilon$.

Now we do the following procedure for i counting from 1 to k. Let (w_i, w'_i) be a factorization of w'_{i-1} such that w_i is the maximal prefix of w'_{i-1} where for all prefix u (if $a_i \neq \epsilon$, u should not be ϵ) of w_i, we have $\gamma(w_0 a_1 \dots a_i u)\mathcal{R}\gamma(w_0 a_1 \dots a_i w_i)$. If w'_i is not left-open, let $w'_i = a_{i+1}w''_i$. Otherwise take $a_i = \epsilon$. Rewrite w'_i with w''_i if $a_i \neq \epsilon$.

The sequence we get from the above procedure $(w_0, a_1, \dots a_k, w_k)$ satisfy all the properties of $\mathcal{R}$ decomposition. ◀

The following Lemma connects $\mathcal{R}$ decompositions and left decompositions.

▶ **Lemma 14.** *Let $(w_0, a_1, \dots, a_k, w_k)$ be the $\mathcal{R}$ decomposition of w. Then, for each $0 < i \leq k$,*

1. *If a_i is not empty, then $a_i \notin alphabet(w_{i-1})$.*
2. *If a_i is empty, then there exists an $a \notin alphabet(w_{i-1})$ such that $a \in T_l^{\omega^*}(w'_i)$ for all nonempty prefix w'_i of w_i.*

Proof. We first need a property to understand when there is no $\mathcal{R}$ class drop.

▶ **claim (⋆).** *Let $x, y, z \in A^\circ$ such that $\gamma(x)\mathcal{R}\gamma(xy)$ and $alphabet(z) \subseteq alphabet(y)$. Then $\gamma(x)\mathcal{R}\gamma(xyz)$.*

Proof. From the assumptions we have a $t \in A^\circ$, such that $\gamma(x) = \gamma(xyt) = \gamma(x(yt)^n)$. From Lemma 11, we know that $\gamma((yt)^n z(yt)^n) = \gamma((yt)^n)$. Therefore $\gamma(x)\mathcal{R}\gamma(xyz)$. ◀

We prove the property for all $0 < i \leq k$. Our proof depends on two cases, if a_i is empty or not.

1. Let a_i be non empty and let us assume for the sake of contradiction $w_{i-1} = xa_iy$, where $x \neq \epsilon$ if w_{i-1} is left open. Then
$$\begin{aligned} \gamma(w_0 \dots a_{i-1}x)\mathcal{R}\gamma(w_0 \dots a_{i-1}xa_iy) & \qquad (\because \text{Only one of } a_{i-1} \text{ and } x \text{ can be empty}) \\ \mathcal{R}\gamma(w_0 \dots a_{i-1}xa_iya_i)\ . & \qquad (\text{From } (\star)) \end{aligned}$$
This is a contradiction.
2. Let a_i be empty, then by definition w_i does not have a minimal point. Let $S = \{b_1, \dots, b_l\} \subseteq T_l^{\omega^*}(w'_i)$ for all non empty prefix w'_i of w_i. We will show that there exists a $b \in S$, such that $b \notin alphabet(w_{i-1})$. Let us assume for the sake of contradiction that $w_{i-1} = x_0 b_1 x_1 \dots b_l x_l$, where $x_i \in A^\circ$ can be empty. Then we have
$$\begin{aligned} \gamma(w_0 \dots a_{i-1}x_0)\mathcal{R}\gamma(w_0 \dots a_{i-1}w_{i-1}) & \qquad (\because \text{Only one of } a_{i-1} \text{ and } x_0 \text{ can be empty}) \\ \mathcal{R}\gamma(w_0 \dots a_{i-1}w_{i-1}w'_i)\ . & \qquad (\text{From } (\star)) \end{aligned}$$
This is a contradiction.

◀

We are now in a position to prove our claim.

Proof of Theorem 8, $(4 \Rightarrow 3)$**.** We show that if $u \equiv_m v$ for a sufficiently large m (depending only on $alphabet(u)$ and $\mathbf{M}$), then $\gamma(u)\mathcal{R}\gamma(v)$. The $\mathcal{L}$ equivalence can be shown symmetrically. As discussed in the beginning, this proves our claim. Our induction hypothesis is as follows:

$$\text{If } u \equiv_m v \text{ for an } m > |alphabet(u)| \times |\mathbf{M}|, \text{ then } \gamma(u) = \gamma(v).$$

The base case, when $m = 0$ is clearly true, since $u = v = \epsilon$ (note that, in this case $alphabet(u) = \emptyset$). Let us now consider the inductive step, for $m > 0$, we have $u \equiv_m v$. Our aim is to show that $\gamma(u) = \gamma(v)$. Consider the $\mathcal{R}$ decomposition of $u = (u_0, a_1, u_1, \dots, a_k, u_k)$.

We give a sequence $v = (v_0, a_1, v_1, \dots, a_k, v_k)$ such that $\gamma(u_i) = \gamma(v_i)$ for all $i < k$ and hence $\gamma(u) \ \geq_{\mathcal{R}} \ \gamma(v)$.

Define $u'_i = u_i a_{i+1} \dots u_k$, for all $i \leq k$. We do the following procedure for i ranging from $1, 2, \dots, k$. During every iteration of i, we give v'_i, a suffix of v_i such that the invariant $u'_i \equiv_{m-i} v'_i$ is maintained. To start the iteration we set $v'_0 = v$ and $u'_0 \equiv_m v'_0$

1. If a_i is non empty, then (u_{i-1}, a_i, u'_i) is the a_i-left decomposition of the word u'_{i-1} (follows from Lemma 14). Since $(u'_{i-1} \equiv_{m-(i-1)} v'_{i-1})$, there exists an a_i-left decomposition of $v'_{i-1} = (v_{i-1}, a_i, v'_i)$ such that $u_{i-1} \equiv_{m-(i-1)} v_{i-1}$ and $u'_i \equiv_{m-i} v'_i$.
2. If a_i is empty, then (u_{i-1}, u'_i) is an $[a]_l$-left decomposition of the word u'_{i-1} for an $[a]_l \in T_l^{\omega^*}(u'_{i-1})/ \sim_l$ (follows from Lemma 14). Since $(u'_{i-1} \equiv_{m-(i-1)} v'_{i-1})$, there exists an $[a]_l$-left decomposition of $v'_{i-1} = (v_{i-1}, v'_i)$ such that $u_{i-1} \equiv_{m-(i-1)} v_{i-1}$ and $u'_i \equiv_{m-i} v'_i$.

Assign $v_k = v'_k$ obtained at the end of iteration.

Note that $k \leq |\mathbf{M}|$. For an $i < k$, we have $|alphabet(u_i)| = |alphabet(v_i)| < |alphabet(u)|$ (from Lemma 14) and therefore $m - i > |alphabet(u_i)| \times |\mathbf{M}|$. Since $u_i \equiv_{m-i} v_i$ from induction hypothesis, it follows $\gamma(u_i) = \gamma(v_i)$, for all $i < k$. Therefore $\gamma(u_0 \dots a_k) = \gamma(v_0 \dots a_k)$.

It remains to show that $\gamma(u) \ \geq_{\mathcal{R}} \ \gamma(v)$. Depending on whether a_k is empty or not, we get the following cases.

1. If a_k is non empty, then $\gamma(u_0 a_1 \dots a_k u_k) \ \mathcal{R} \ \gamma(u_0 a_1 \dots a_k) = \gamma(v_0 a_1 \dots a_k) \geq_{\mathcal{R}} \gamma(v)$. The first condition follows from the fact that the sequence $(u_0 a_1 \dots u_k)$ is an $\mathcal{R}$ decomposition, and the second condition follows from the fact that $\gamma(u_i) = \gamma(v_i)$ for all $i < k$.
2. If a_k is empty, then $(u_0 \dots u_{k-1}, u_k)$ and $(v_0 \dots v_{k-1}, v_k)$ are both S-left decomposition for an $S \in T_l^{\omega^*}(u_i)/ \sim_l$. Hence there are prefixes u'_k of u_k and v'_k of v_k such that $u'_k, v'_k \in S^{-\infty}$. From Lemma 11 we know that $\gamma(u'_k) \mathcal{R} \gamma(v'_k)$. Therefore, $\gamma(u_0 a_1 \dots a_k u_k) \ \mathcal{R} \ \gamma(u_0 a_1 \dots u'_k) \ \mathcal{R} \ \gamma(v_0 a_1 \dots v'_k) \ \ \geq_{\mathcal{R}} \ \ \gamma(v_0 a_1 \dots a_k v_k) = \gamma(v)$.

We now have $\gamma(u) \geq_{\mathcal{R}} \gamma(v)$. By a symmetric argument we get $\gamma(v) \geq_{\mathcal{R}} \gamma(u)$ and therefore $\gamma(u) \ \mathcal{R} \ \gamma(v)$. By $\mathcal{L}$-$\mathcal{R}$ symmetry, $\gamma(u) \ \mathcal{L} \ \gamma(v)$ and since $\mathbf{M}$ is aperiodic $\gamma(u) = \gamma(v)$. ◀

3.4 Logic to Algebra

In this subsection we show direction (2 ⇒ 4) of Theorem 8. We will show that, if L is definable by an FO^2(<) sentence, then L is recognizable by a ∘-DA. It suffices to show that the game congruence $\cong$ satisfies the equations of ∘-DA.

We will show that for all numbers $r > n$ the duplicator wins the an r round EF game on the different equations. Our proof is a case analysis for the different equations.

- The equation $(xyz)^n y (xyz)^n = (xyz)^n$: The proof that duplicator wins the r round EF game can be found in [26].
- The equation $(x^n)^\omega (x^n)^{\omega^*} = x^n$: Follows, since FO satisfies this equation [6].
- The equation $\{x_1, \dots, x_n\}^\eta = (x_1 \dots x_n)^{\omega^*} (x_1 \dots x_n)^\omega$: In the first round, lets say Spoiler pebbles some x_i on one of the structures. Duplicator pebbles an x_i on the other structure. In the subsequent rounds, let Spolier pebble some x_j on one of the structures. Duplicator will pebble an x_j on the other structure such that it preserves the ordering of the two pebbles in both the structures. This is always possible, since both the structures are both left and right infinite.

4 Satisfiability

In this section we address the satisfiability problem of two-variable logic over countable linear orderings. The rest of the section is devoted to the proof of the below theorem. Take note of the fact that in this section Σ denotes a set of unary predicates (and not an alphabet). Our models are words over the alphabet $\mathcal{P}(\Sigma)$.

▶ **Theorem 15.** *The following problems are* Nexptime*-complete: Satisfiability of* $\mathrm{FO}^2(\Sigma, <)$ *over*

1. *arbitrary linear orderings,*
2. *countable linear orderings,*
3. *scattered linear orderings.*

First we deal with the hardness part of the theorem. By downward Löwenheim-Skolem theorem, every satisfiable first-order formula has a countable model, and therefore (1) reduces to (2). Similary by Lemma 21 (given below), if a two-variable logic formula has a countable model, then it has a scattered model. Therefore (2) reduces to (3). Secondly, satisfiability of $\mathrm{FO}^2(\Sigma)$ over arbitrary structures already is Nexptime-hard [9], and therefore (1), (2) and (3) are Nexptime-hard.

Next we prove that (2) and (3) are in Nexptime. The idea is to show that for any satisfiable formula there is a model of a particular form that admit at most exponentially big (in the size of the formula) description.

Let φ be a $\mathrm{FO}^2(\Sigma, <)$ formula. Using standard ideas we obtain a formula $\varphi' \in \mathrm{FO}^2(\Sigma', <)$ in Scott normal form, i.e.

$$\varphi' = \forall x \forall y\, \psi(x,y) \wedge \bigwedge_i \forall x \exists y\, \chi_i(x,y)\,, \tag{1}$$

where $\Sigma' \supseteq \Sigma$, $|\Sigma'| = |\Sigma| + \mathcal{O}(|\varphi|)$, $|\varphi'| = \mathcal{O}(|\varphi|)$, $\psi(x,y)$ and $\chi_i(x,y)$ are quantifier free, such that φ and φ' are equisatisfiable (one is satisfiable if and only if the other is satisfiable). More precisely, the sets of models of φ and φ' are isomorphic upto the erasure of the unary predicates $\Sigma' \setminus \Sigma$.

We introduce some notation. Given a set of unary predicates P, we define a unary type over P to be a maximal conjunction of literals (i.e. $U(x)$ or $\neg U(x)$ where U is a unary predicate in P) over the same variable that is satisfiable. When the set P is clear from the context we just use types to refer to the unary types over P. We write $\mathrm{tp}(P)$ to denote the types over the predicates P. Each position of a $\circ$-word satisfies exactly one type, called the *type of the position*. Models of φ' are $\circ$-words over the alphabet $\mathrm{tp}(\Sigma')$.

▶ **Lemma 16.** *Each formula* $\forall x \exists y\, \chi_i(x,y)$ *is equivalent to a formula*

$$\bigwedge_j \forall x \left(\alpha_j(x) \rightarrow \exists y \left(O(x,y) \wedge \bigvee_k \beta_{jk}(y) \right) \right) \tag{2}$$

where α_j, β_{jk} *are types and* $O(x,y)$ *is a disjunction over the set* $\{x < y, x = y, x > y\}$.

Proof. By writing each $\chi_i(x,y)$ in disjunctive normal form, the formula $\forall x \exists y\, \chi_i(x,y)$ is equivalent to a formula of the form

$$\forall x \exists y \bigvee_l \alpha_l(x) \wedge O_l(x,y) \wedge \beta_l(y) \tag{3}$$

where $\alpha_l(x)$ and $\beta_l(y)$ are a conjunctions of literals over the variable x and y respectively and $O_l(x, y)$ is a disjunction over the set $O = \{x < y, x = y, x > y\}$. By adding literals to each α_l and β_l, it can be seen that formula 3 is equivalent to

$$\forall x \exists y \bigvee_j \left(\alpha_j(x) \wedge O_j(x, y) \wedge \bigvee_k \beta_{jk}(y) \right) \tag{4}$$

where each α_i and each β_{ij} are unary types and $O_j(x, y)$ is a disjunction over O. Moreover, since each position satisfies exactly one α_j the Formula 4 is equivalent to

$$\forall x \exists y \bigwedge_j \left(\alpha_j(x) \to O_j(x, y) \wedge \bigvee_k \beta_{jk}(y) \right) \ . \tag{5}$$

Finally we observe that this is equivalent to

$$\bigwedge_j \forall x \left(\alpha_j(x) \to \exists y \left(O_j(x, y) \wedge \bigvee_k \beta_{jk}(y) \right) \right) \ . \tag{6}$$

◀

Next we show that given any formula in Scott normal form it has models of a particular form. Before that we need some elementary lemmas about linear orderings.

▶ **Lemma 17.** *If $X_1, \ldots, X_n \subseteq Z$ are right-open sets such that for each $1 \leq i < n$, X_{i+1} contains an upperbound of X_i, then there exist nonempty suffixes $X'_1, \ldots, X'_n$ of $X_1, \ldots, X_n$ respectively such that $X'_1 < \cdots < X'_n$.*

Proof. Assume that $X_1, \ldots, X_n$ are right-open subsets of Z and for each $1 \leq i < n$, X_{i+1} contains an upperbound of X_i. Choose $x_2 \in X_2, \ldots, x_n \in X_n$ such that for each $2 \leq i \leq n$, the element x_i is an upperbound of X_{i-1}. For $2 \leq i \leq n$, let $X'_i = \{y \in X_i : y > x_i\}$. Then, the sets $X_1 < X'_2 < \cdots < X'_n$ are nonempty suffixes of $X_1, \ldots, X_n$ respectively. ◀

Dualy the following also holds.

▶ **Lemma 18.** *If $X_1, \ldots, X_n \subseteq Z$ are left-open sets such that for each $1 \leq i < n$, X_i contains a lowerbound of X_{i+1}, then there exist nonempty prefixes $X'_1, \ldots, X'_n$ of $X_1, \ldots, X_n$ respectively such that $X'_1 < \cdots < X'_n$.*

▶ **Lemma 19.** *If $\bar{X} = X_1 < \cdots < X_m$ is a sequence of left-open sets and $\bar{Y} = Y_1 < \cdots < Y_n$ is a sequence of right-open sets then there exist $X'_1 < \cdots < X'_m$ and $Y'_1 < \cdots < Y'_n$ such that*

1. *for each $1 \leq i \leq m$, X'_i is a nonempty prefix of X_i,*
2. *for each $1 \leq i \leq n$, Y'_i is a nonempty suffix of Y_i, and*
3. *for each pair $1 \leq i \leq m, 1 \leq j \leq n$, either $X'_i < Y'_j$ or $Y'_j < X'_i$, i.e. the set $\{X'_1, \ldots, X'_m, Y'_1, \ldots, Y'_n\}$ is linearly ordered by the relation $<$.*

Proof. We prove the claim by induction on the set of all pairs $(\bar{X} = X_1 < \cdots < X_m, \bar{Y} = Y_1 < \cdots < Y_n)$ of sequences of nonempty sets — such that X_i's are left-open, and Yi's are right-open — ordered pointwise by the prefix ordering on sequences. The induction base is the degenerate case when both $\bar{X}$ and $\bar{Y}$ are empty sequences, and the claim holds vacuously. For the inductive step assume that the claim holds $(\bar{X} = X_1 < \cdots < X_m, \bar{Y} = Y_1 < \cdots < Y_n)$. We have two cases to consider, namely the pairs $(X_1 < \cdots < X_m < X, \bar{Y})$ and $(\bar{X}, Y_1 < \cdots < Y_n < Y)$ where X and Y are some nonempty subsets of Z such that

X is left-open and Y is right-open. Next we prove the claim for both cases. By induction hypothesis, there exist sequences $\bar{X}' = X'_1 < \cdots < X'_m$ and $\bar{Y}' = Y'_1 < \cdots < Y'_n$ that satisfy conditions (1), (2) and (3).

Consider the pair $(X_1 < \cdots < X_m < X, \bar{Y})$. If X does not intersect with any of the sets $Y'_1, \ldots, Y'_n$, then the sequences $\bar{X}'_1 < \cdots < X'_m < X$ and $\bar{Y}'$ satisfies the claim. Otherwise, let $1 \leq j \leq n$ be the smallest index such that $Y'_j \cap X \neq \emptyset$. Choose an $z \in Y'_j \cap X$ and define $X' = \{x \in X : x < z\}$ and $Y''_j = \{y \in Y'_j : y > z\}$. We verify that the sequences

$$X'_1 < \cdots < X'_m < X' \text{ and } Y'_1 < \cdots < Y'_{j-1} < Y''_j < Y'_{j+1} < \cdots < Y'_n$$

satisfy the claim. Clearly $X' < Y''_j$. Also, for each $1 \leq i < j$, $Y'_j < X'$ and for each $j < i \leq m$, $X' < Y'_i$. All other cases follow from the induction hypothesis. Thus the claim is verified.

Next consider the pair $(\bar{X}, Y_1 < \cdots < Y_n < Y)$. If Y does not intersect with any of the X'_i then $\bar{X}'$ and $Y'_1 < \cdots < Y'_n < Y$ satisfies the claim. Otherwise we choose the largest index $1 \leq i \leq n$ such that $X'_i \cap Y$ is nonempty and choose an $z \in X'_i \cap Y$. Let $Y' = \{y \in Y : y > z\}$ and $X''_j = \{x \in X'_j : x < z\}$. We claim that the sequences

$$X'_1 < \cdots < X'_{j-1} < X''_j < X'_{j+1} < \cdots < X'_m \text{ and } Y'_1 < \cdots < Y'_n < Y'$$

satisfy the claim. By definition, $X''_j < Y'$. For each $1 \leq j < i$, $X'_j < Y'$, and for each $i < j \leq n$, $Y' < X'_j$. Rest of the cases are satisfied by the induction hypothesis. Thus the claim is verified. ◀

▶ **Lemma 20.** *If $X \subseteq Z$ is finite and $Y \subseteq Z$ is right-open (resp. left-open) then there exist disjoints subsets X_1, X_2 of X and a nonempty suffix (resp. prefix) Y' of Y such that $X_1 < Y' < X_2$ and $X = X_1 \cup X_2$.*

Proof. The claims are dual and we treat only one case. Assume $X \subseteq Z$ is finite and $Y \subseteq Z$ is right-open. If $Y < X$ then $\emptyset < Y < X$ and the claim follows. Otherwise let $x \in X$ be the maximal element in X for which there is some $y \in Y$ such that $x < y$. We take $X_1 = \{y \in X : y \leq x\}$ and observe that $X_1 < Y' = \{y \in Y : y > x\} < X \setminus X_1$ and the set Y' is a nonempty suffix of Y. ◀

Next we prove that formulas φ' in Scott normal form possess particular kind of models.

▶ **Lemma 21.** *If φ' is satisfiable, then it has a model of the form $u_1^{\lambda_1} \cdots u_n^{\lambda_n}$ where $n \geq 1$ is a natural number, for each $1 \leq i \leq n$, u_i is a finite word over the alphabet $\mathrm{tp}(\Sigma')$ and λ_i is in $\{1, \omega, \omega^*\}$, such that*

1. *every type occurs at most once in each u_i, and*
2. *every type occurs in at most two u_i's.*

Proof. Assume φ' is satisfiable and let u be a $\circ$-word over the alphabet $\mathrm{tp}(\Sigma')$ that satisfies it. Let T be the set of all types occurring in u. For a type α in T, let $P(\alpha)$ denote the set of all positions in u labelled with α. Let $T_r^1 \subseteq T$ be the set of all types α such that $P(\alpha)$ has a maximal element. Further more let T_r^ω be the set $T \setminus T_r^1$.

We define a total preorder $\lesssim_r$ over the set of types T_r^ω as follows.

$\alpha \lesssim_r \beta$ if each α-position in u has a β-position to its right.

We verify that $\lesssim_r$ is indeed a total preorder. Since for each type α in T_r^ω the set $P(\alpha)$ does not have a maximum, clearly $\alpha \lesssim_r \alpha$. Next, let $\alpha, \beta, \gamma \in T_r^\omega$ be such that $\alpha \lesssim_r \beta$ and $\beta \lesssim_r \gamma$. By definition of $\lesssim_r$, every α-position has a β-position to the right, which in turn has a γ-position to its right. Hence $\alpha \lesssim_r \gamma$. It only remains to show that $\lesssim_r$ is total. Assume $\alpha \not\lesssim_r \beta$. By definition, there is an α-position i that has no β-position to its right. Hence all β-positions lie to the left of i, and therefore every β-position has an α to its right, and therefore $\beta \lesssim_r \alpha$. We write $\sim_r$ to denote the equivalence relation associated with the preorder $\lesssim_r$. For a type α in T_r^ω we let $[\alpha]_r \subseteq T_r^\omega$ denote the equivalence class of α with respect to the total preorder $\lesssim_r$, i.e. $[\alpha]_r = \{\beta \in T_r^\omega : \beta \sim_r \alpha\}$. We write $<_r$ to denote the total order on $\{[\alpha]_r : \alpha \in T_r^\omega\}$.

Let $[\alpha_1]_r <_r \cdots <_r [\alpha_h]_r$ be the classes of the equivalence $\sim_r$. For a class $[\alpha_i]_r$, we define $P([\alpha_i]_r)$ to be the set $\cup_{\beta \sim_r \alpha_i, \beta \in T_r^\omega} P(\beta)$. By definition of the preorder $\lesssim_r$, each $P([\alpha_i]_r)$ has an upperbound in $P([\alpha_{i+1}]_r)$. Moreover, none of the $P([\alpha_i]_r)$ has a maximal element. Applying the Lemma 17 to the sets $P([\alpha_1]_r), \ldots, P([\alpha_h]_r)$ we obtain their respective suffixes $Q_{[\alpha_1]_r}, \ldots, Q_{[\alpha_h]_r}$ guaranteed by the lemma such that $Q_{[\alpha_1]_r} < \cdots < Q_{[\alpha_h]_r}$.

Next we repeat the above construction for the opposite direction. Let $T_l^1 \subseteq T$ is the set of types α such that $P(\alpha)$ has a minimum and let $T_l^{\omega^*} = T \setminus T_l^1$. The dual relation $\lesssim_l$ defined as

$\beta \lesssim_l \alpha$ if each α-position in u has a β-position to its left

is also a total preorder. The corresponding equivalence relation and strict order relation are denoted as $\sim_l$ and $<_l$. Let $[\beta_1]_l <_l \cdots <_l [\beta_k]_l$ be the classes of the equivalence $\sim_l$. As before we let $P([\beta_i]_l)$ to be the set $\cup_{\alpha \sim_l \beta_i, \alpha \in T_l^{\omega^*}} P(\alpha)$. By definition of $\lesssim_l$, each $P([\beta_i]_l)$ contains a lower bound of $P([\beta_{i+1}]_l)$. Also, none of $P([\beta_i]_l)$ has a minimum. Therefore applying Lemma 18 to the sets $P([\beta_1]_l), \ldots, P([\beta_k]_l)$ we obtain prefixes $Q_{[\beta_1]_l}, \ldots, Q_{[\beta_k]_l}$ of $P([\beta_1]_l), \ldots, P([\beta_k]_l)$ such that $Q_{[\beta_1]_l} < \cdots < Q_{[\beta_k]_l}$.

Next we apply the Lemma 19 to the sequences $Q_{[\alpha_1]_r} < \cdots < Q_{[\alpha_h]_r}$, $Q_{[\beta_1]_l} < \cdots < Q_{[\beta_k]_l}$ and obtain suffixes $Q'_{[\alpha_1]_r} < \cdots < Q'_{[\alpha_h]_r}$ and prefixes $Q'_{[\beta_1]_l} < \cdots < Q'_{[\beta_k]_l}$ as in the statement of the lemma.

Let F be the set of positions containing maximal occurrences of types in T_r^1 and minimal occurrences of types in T_l^1, i.e. $F = \bigcup_{\alpha \in T_r^1}\{max(P(\alpha))\} \cup \bigcup_{\alpha \in T_l^1}\{min(P(\alpha))\}$. By applying the Lemma 20 with the finite set F and each of the set $Q'_{[\alpha_1]_r}, \ldots, Q'_{[\alpha_h]_r}$ (which don't have maximums) we obtain their respective suffixes $Q''_{[\alpha_1]_r}, \ldots, Q''_{[\alpha_h]_r}$ such that for each $Q''_{[\alpha_i]}$ there exists $X, Y \subseteq F$ such that $X < Q''_{[\alpha_i]_r} < Y$ and $X \cup Y = F$. Similarly again applying the Lemma 20 with H and each of the set $Q'_{[\beta_1]_l}, \ldots, Q'_{[\beta_k]_l}$ (which don't have minimums) we obtain their respective prefixes $Q''_{[\beta_1]_r}, \ldots, Q''_{[\beta_k]_l}$ such that for each $Q''_{[\beta_i]}$ there exists $X, Y \subseteq F$ such that $X < Q''_{[\beta_i]_l} < Y$ and $X \cup Y = F$. Therefore we conclude that there exists nonempty disjoint subsets $F_1, \ldots, F_t \subseteq F$ such that $F = \cup_{i=1}^t F_t$ and the set $\{Q''_{[\alpha_1]_r}, \ldots, Q''_{[\alpha_h]_r}, Q''_{[\beta_1]_l}, \cdots, Q''_{[\beta_k]_l}, F_1, \ldots, F_t\}$ is linearly ordered by the relation $<$. Finally from each set $Q''_{[\alpha_i]_r}$ where $[\alpha_i]_r = \{\alpha_1, \ldots, \alpha_k\}$ we choose a countable set of positions $I_{[\alpha_i]_r} = \{i_1 < i_2 < \cdots\}$ such that the set $I_{[\alpha_i]_r}$ constitutes the ω-word $(\alpha_1 \cdots \alpha_k)^\omega$. Similarly, from each set $Q''_{[\beta_i]_l}$ where $[\beta_i]_r = \{\beta_1, \ldots, \beta_k\}$ we choose a countable set of positions $I_{[\beta_i]_l} = \{i_1 > i_2 > \cdots\}$ such that the set $I_{[\beta_i]_l}$ constitutes the ω^*-word $(\beta_1 \cdots \beta_k)^{\omega^*}$. We define u' to be the subword of u (our initial model) given by the set of positions

$$I = \bigcup_{i=1}^{h} I_{[\alpha_i]_r} \cup \bigcup_{i=1}^{k} I_{[\beta_i]_l} \cup \bigcup_{i=1}^{t} F_i$$

and show that u' satisfies the conditions described by the Lemma. By definition of the sets $I_{[\alpha_i]_r}$, $1 \leq i \leq h$, and $I_{[\beta_i]_l}$, $1 \leq j \leq k$, the $\circ$-word u' is of the form $u_1^{\lambda_1} \cdots u_n^{\lambda_n}$ where u_i

are finite words over the alphabet T and λ_i are in $\{1, \omega, \omega^*\}$. Conditions (1) and (2) of the lemma simply follows by construction.

It only remains to show that u' is also a model of the formula. Firstly we observe that since u' is a sub-$\circ$-word of u it satisfy the formula $\forall x \forall y\, \psi(x, y)$ from the Equation 1. Next we need to show that u' satisfies each of the formula $\zeta = \forall x \exists y\, \chi_i(x, y)$. By Lemma 16 ζ is equivalent to a formula of the form $\bigwedge_j \forall x\ (\alpha_j(x) \to \exists y\ (O_j(x, y) \wedge (\vee_j \beta_{jk}(y)))$ where α_j, β_{jk} are types and $O_j(x, y)$ is a disjunction over $\{x < y, x = y, x > y\}$. Therefore, to show that u' satisfies the formula ζ it is enough to show that if an α-position in I has some β-position y to the right (*resp.* left), i.e. $y > x$, in u, then it has a β-position to the right (*resp.* left) in u' also, i.e there is some y' in I such that $y' > x$. We only consider the case when β occurs to the right. We do a case analysis. If $P(\beta)$ has a maximum element then that is a witness for x, and it is present in u'. Otherwise the set $P(\beta)$ is right-open and the set $I_{[\beta]_r}$ contains a suffix of $P([\beta])$ by construction. Hence there is β to the right of α. ◀

A model of the form $u = u_1^{\lambda_1} \cdots u_n^{\lambda_n}$ is finitely represented as a sequence of pairs $(u_1, \lambda_1) \cdots (u_n, \lambda_n)$. Lemma 21 guarantees that for every satisfiable formula φ' there is a representation of size at most $3 \cdot \mathrm{tp}(\Sigma) \le 3 \cdot 2^{|\varphi'|}$.

▶ **Lemma 22.** *Given a sequence of pairs $(u_1, \lambda_1) \cdots (u_n, \lambda_n)$ and a formula φ' checking if the $\circ$-word $u_1^{\lambda_1} \cdots u_n^{\lambda_n}$ satisfies the formula φ in Scott normal form is in* Ptime.

Proof. First we prove a small claim. Let A be an alphabet and let v be a finite word over A. Let $v_1, v_2, v_3, v_1', v_2', \ldots$ be copies of the word v. There is an obvious correspondence between positions of any two copies, that maps the ith position of one copy to the ith position of the other copy. Let $u, u', w, w' \in A^\circ$ be words such that $alphabet(u) = alphabet(u')$ and $alphabet(w) = alphabet(w')$. We define B and C to be the words

$$B = u \cdot v_1 \cdot v_2 \cdot v_3 \cdot w \qquad\qquad C = u' \cdot v_1' \cdot v_2' \cdots \cdot w' \ .$$

We claim that ($\star$) for positions x in B and y in C, the pairs $B, x \cong_1 C, y$ whenever (1) x and y are corresponding positions from v_1 and v_1' respectively, or (2) x and y are corresponding positions from copies v_2 and v_j', $j \ge 2$ respectively. To prove the claim observe that starting from any configuration where x and y are selected in the words B and C, the duplicator has a winning strategy in the 1-round game, whenever x and y satisfies one of the previous conditions: whenever the spoiler picks an a-position to the left (*resp.* right) she also picks an a-position to the left (*resp.* right). This is always possible since the letters on the left (as well as right) of x and y are the same. Therefore $B, x \equiv_1 C, y$, i.e. they satisfy the same first order formulas with a free variable with quantifier rank at most 1. Similarly the dual claim also holds: If B and C are respectively the words

$$B = u \cdot v_3 \cdot v_2 \cdot v_1 \cdot w \qquad\qquad C = u' \cdot \ \cdots v_2' \cdot v_1' \cdot w' \ .$$

then for positions x in B and y in C, the pairs $B, x \cong_1 C, y$ whenever Conditions (1) and (2) are met.

Next we prove the Lemma. Assume that we are given a sequence of pairs $\bar{u} = (u_1, \lambda_1) \cdots (u_n, \lambda_n)$ and a formula $\varphi = \forall x \forall y\, \psi(x, y) \wedge \bigwedge_i \forall x \exists y \chi_i(x, y)$. Let $u \in A^\circ$ be the word $u_1^{\lambda_1} \cdots u_n^{\lambda_n}$ and let u' be the finite word $u_1^{\lambda_1'} \cdots u_n^{\lambda_n'}$ where λ_i' is 1 if λ_i is 1, and 3 otherwise. The subset of positions of u' that correspond to the case $\lambda_i = 1$ is called *finitary* positions. The word u satisfies the formula φ if every position of u satisfies the set of formulas $S = \{\forall y\, \psi(x, y), (\exists y\, \chi_i(x, y))_i\}$. By Claim ($\star$) and its dual, every position in a factor of u of the form u_i^ω or $u_i^{\omega^*}$ is $\equiv_1$-equivalent to a position in u_i^3, namely the positions in the first two copies of u_i, or the last two copies of u_i respectively. Therefore it is enough to verify the set S on such positions, as well as on the finitary positions. This can be done in time $\mathcal{O}\left(|\varphi|^2 \cdot |\bar{u}|^2\right)$.

◀

To complete the proof of the Theorem 15 we describe a NEXPTIME algorithm for FO^2 formulas over countable linear orders: The algorithm converts the input formula to Scott normal form and guesses an atmost exponentially large representation of a model of the form described by Lemma 21 and checks that it is indeed a model by Lemma 22.

5 Conclusion

In this paper we characterised first-order logic with two variables over countable linear orderings. It is equivalent to a fragment of temporal logic and is characterised by a subclass of $\circ$-monoids, called $\circ$-DA. The class $\circ$-DA is the class of $\circ$-monoids whose regular $\mathcal{J}$ classes are sub $\circ$-monoids. We also proved an alternate characterisation of this class using equations and this yields decidability of membership in this class. Next we considered the satisfiability problem for FO^2 over arbitrary, countable and scattered linear orderings and showed that all the problems are NEXPTIME-complete.

Finally we note that FO^2 with order and *successor* relation (position $j > i$ is the successor of position i if there is no position between them) is strictly more powerful that FO^2 with only the order relation. To see this it is enough to note that a^ω and $a^\omega a^\omega$ are indistinguishable by any formula in the latter class, while there is a formula, namely "there is exactly one position without a predecessor" that separates them. We leave as future work the question of extending the characterisation in the present paper to handle the successor relation.

References

1 Nicolas Bedon, Alexis Bès, Olivier Carton, and Chloe Rispal. Logic and rational languages of words indexed by linear orderings. *Theory Comput. Syst.*, 46(4):737–760, 2010.

2 Alexis Bès and Olivier Carton. Algebraic characterization of FO for scattered linear orderings. In *Computer Science Logic, 25th International Workshop / 20th Annual Conference of the EACSL, CSL 2011*, pages 67–81, 2011.

3 Mikołaj Bojańczyk, Claire David, Anca Muscholl, Thomas Schwentick, and Luc Segoufin. Two-variable logic on data words. *ACM Trans. Comput. Log.*, 12(4):27, 2011.

4 Olivier Carton, Thomas Colcombet, and Gabriele Puppis. Regular languages of words over countable linear orderings. In *Automata, Languages and Programming - 38th International Colloquium, ICALP 2011, Proceedings, Part II*, pages 125–136, 2011.

5 Thomas Colcombet. Factorization forests for infinite words and applications to countable scattered linear orderings. *Theor. Comput. Sci.*, 411(4-5):751–764, 2010.

6 Thomas Colcombet and A. V. Sreejith. Limited set quantifiers over countable linear orderings. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Proceedings, Part II*, pages 146–158, 2015.

7 Julien Cristau. Automata and temporal logic over arbitrary linear time. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2009*, pages 133–144, 2009.

8 Kousha Etessami, Moshe Y. Vardi, and Thomas Wilke. First-order logic with two variables and unary temporal logic. *Inf. Comput.*, 179(2):279–295, 2002.

9 Martin Fürer. The computational complexity of the unconstrained limited domino problem (with implications for logical decision problems). In *Logic and Machines: Decision Problems and Complexity, Proceedings of Symposium Rekursive Kombinatorik*, pages 312–319, 1983.

10 Erich Grädel, Phokion G. Kolaitis, and Moshe Y. Vardi. On the decision problem for two-variable first-order logic. *Bulletin of Symbolic Logic*, 3(1):53–69, 1997.

11 Manfred Kufleitner and Pascal Weil. On FO2 quantifier alternation over words. In *Mathematical Foundations of Computer Science 2009, 34th International Symposium, MFCS 2009*, pages 513–524, 2009.

12 Manfred Kufleitner and Pascal Weil. The FO2 alternation hierarchy is decidable. In *Computer Science Logic (CSL'12) - 26th International Workshop/21st Annual Conference of the EACSL, CSL 2012*, pages 426–439, 2012.

13 Amaldev Manuel. Two variables and two successors. In *Mathematical Foundations of Computer Science 2010, 35th International Symposium, MFCS*, pages 513–524, 2010.

14 Amaldev Manuel and Thomas Zeume. Two-variable logic on 2-dimensional structures. In *Computer Science Logic 2013 (CSL 2013), CSL*, pages 484–499, 2013.

15 Martin Otto. Two variable first-order logic over ordered domains. *J. Symb. Log.*, 66(2):685–702, 2001.

16 Jean-Éric Pin. Mathematical foundations of automata theory.

17 Jean-Eric Pin and Pascal Weil. Polynomial closure and unambiguous product. In *Automata, Languages and Programming, 22nd International Colloquium, ICALP95, Proceedings*, pages 348–359, 1995.

18 Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.

19 Alexander Rabinovich. Temporal logics over linear time domains are in PSPACE. *Inf. Comput.*, 210:40–67, 2012.

20 Joseph G. Rosenstein. *Linear orderings.* Academic Press New York, 1981.

21 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.

22 Thomas Schwentick, Denis Thérien, and Heribert Vollmer. Partially-ordered two-way automata: A new characterization of DA. In *Developments in Language Theory, 5th International Conference, DLT 2001*, pages 239–250, 2001.

23 Thomas Schwentick and Thomas Zeume. Two-variable logic with two order relations. *Logical Methods in Computer Science*, 8(1), 2012.

24 S.Shelah. The monadic theory of order. *Ann. of Math.*, 102:379–419, 1975.

25 Pascal Tesson and Denis Therien. Diamonds are forever: The variety DA. In *Semigroups, Algorithms, Automata and Languages*, pages 475–500. World Scientific, 2002.

26 Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 234–240. ACM, 1998.

27 Philipp Weis and Neil Immerman. Structure theorem and strict alternation hierarchy for FOˆ2 on words. *Logical Methods in Computer Science*, 5(3), 2009.