

On the decidability of MSO+U on infinite trees

Mikołaj Bojańczyk¹, Tomasz Gogacz², Henryk Michalewski¹ and Michał Skrzypczak^{1*}.

¹ University of Warsaw

² University of Wrocław

Abstract. This paper is about MSO+U, an extension of monadic second-order logic, which has a quantifier that can express that a property of sets is true for arbitrarily large sets. We conjecture that the MSO+U theory of the complete binary tree is undecidable. We prove a weaker statement: there is no algorithm which decides this theory and has a correctness proof in ZFC. This is because the theory is undecidable, under a set-theoretic assumption consistent with ZFC, namely that there exists of projective well-ordering of 2^ω of type ω_1 . We use Shelah’s undecidability proof of the MSO theory of the real numbers.

1 Introduction

This paper is about MSO+U, which is the extension of MSO by the *unbounding quantifier*. The unbounding quantifier, denoted by

$$\text{UX. } \varphi(X),$$

says that $\varphi(X)$ holds for arbitrarily large finite sets X . As usual with quantifiers, the formula $\varphi(X)$ might have other free variables except for X . The main contribution of the paper is the following theorem, which talks about the complete binary tree 2^* .

Theorem 1.1. *Assuming that there exists a projective well-ordering of 2^ω of type ω_1 , it is undecidable if a given sentence of MSO+U is true in the complete binary tree.*

The assumption on the projective ordering can be seen as a set theory axiom. The assumption follows from the axiom $V=L$, which is relatively consistent with ZFC. Therefore, if ZFC has a model, then it has one where the assumption of Theorem 1.1 is true, and therefore it has a model where the MSO+U theory of the complete binary tree is undecidable. In particular, there is no algorithm which decides the MSO+U theory of the complete binary tree, and has a correctness proof in ZFC. Although the theorem stops short of full undecidability, which we conjecture to be the case, it seems to settle the decidability question for all practical purposes.

* The first and fourth author are supported by ERC Starting Grant “Sosna”, the third author is supported by the Polish NCN grant DEC-2012/07/D/ST6/02443.

Background. This paper is part of a programme researching the logic $\text{MSO}+\text{U}$, i.e. monadic second-order logic extended with the U quantifier. The logic was introduced in [Boj04], where it was shown that satisfiability is decidable for formulae on infinite trees where the U quantifier is used once and not under the scope of set quantification. A significantly more powerful fragment of the logic, albeit for infinite words, was shown decidable in [BC06] using automata with counters. These automata were further developed into the theory of cost functions initiated by Colcombet in [Col09]. Cost functions can be seen as a special case of $\text{MSO}+\text{U}$ in the sense that decision problems regarding cost functions, such as limitedness or domination, can be easily encoded into satisfiability of $\text{MSO}+\text{U}$ formulae. This encoding need not be helpful, since the unsolved problems for cost functions get encoded into unsolved problems from $\text{MSO}+\text{U}$.

The added expressive power of $\text{MSO}+\text{U}$ can be used to solve problems that do not have a simple solution in MSO alone. An example is the star height problem, one of the most difficult problems in the theory of automata, which can be straightforwardly reduced to the satisfiability of $\text{MSO}+\text{U}$ on infinite words; the particular fragment of $\text{MSO}+\text{U}$ used in this reduction is decidable by [BC06]. An example of an important unsolved problem that reduces to $\text{MSO}+\text{U}$ is the decidability of the nondeterministic parity index problem [CL08].

So far, most research on $\text{MSO}+\text{U}$ has focussed on the weak variant, call it $\text{WMSO}+\text{U}$, where only quantification over finite sets is allowed. Satisfiability is decidable for $\text{WMSO}+\text{U}$ over infinite words [Boj11] and infinite trees [BT12]. In a parallel submission to this conference, it is shown that $\text{WMSO}+\text{U}$ remains decidable over infinite trees even after adding quantification over infinite paths. The decidability proofs use automata with counters.

Undecidability. The first strong evidence that $\text{MSO}+\text{U}$ can be too expressive was given in [HS12], where it was shown that $\text{MSO}+\text{U}$ can define languages of infinite words that are arbitrarily high in the projective hierarchy from descriptive set theory. The present paper builds on that observation. We show that, using the languages from [HS12], one can use $\text{MSO}+\text{U}$ on the complete binary tree 2^* to simulate a variant of MSO on the Cantor set 2^ω , which we call *projective* MSO . Projective MSO is like MSO , except that set quantification is restricted to projective sets. As shown by Shelah in [She75], the MSO theory of 2^ω is undecidable. From the proof of Shelah it follows that, under the assumption that there exists a projective well-ordering of 2^ω , already projective MSO is undecidable on 2^ω . Therefore, thanks to our reduction, $\text{MSO}+\text{U}$ is undecidable on 2^* .

2 $\text{MSO}+\text{U}$ on 2^*

We consider the following logical structures: the complete binary tree 2^* , the Cantor set 2^ω , and the union of the two $2^{\leq\omega}$. In the complete binary tree 2^* , the universe consists of finite strings over $\{0,1\}$, called *nodes*, and there are predicates for the lexicographic and prefix orders. The prefix order corresponds to the ancestor relation. In the Cantor set 2^ω , the universe consists of infinite

strings over $\{0, 1\}$, called *branches*, and there is a predicate for the lexicographic order. Finally, in $2^{\leq\omega}$, the universe consists of both nodes and branches, and there are predicates for the prefix and lexicographic order. In $2^{\leq\omega}$, the prefix relation can hold between two nodes, or between a node and a branch. The lexicographic order is a total order on both nodes and branches, e.g. $0 < 0^\omega < 01$.

Two fundamental theorems about MSO are that the MSO theory is decidable for 2^* , but undecidable for 2^ω , and therefore also undecidable for $2^{\leq\omega}$. The decidability was shown by Rabin in [Rab69], while the undecidability was shown by Shelah in [She75] conditionally on the Continuum Hypothesis, and by Shelah and Gurevich in [GS82] without any conditions.

The projective hierarchy. Consider a topological space X . The family of Borel sets is the least family of subsets of X that contains open sets, and is closed under complements and countable unions. Define the family of *projective sets* to be the least family of subsets of X which contains the Borel sets, and is closed under complements and images under continuous functions. The projective sets can be organised into a hierarchy, called the *projective hierarchy*, where $\Sigma_0^1 = \Pi_0^1$ is the class of Borel sets, Π_n^1 is the class of complements of sets from Σ_n^1 , and Σ_{n+1}^1 is the class of images of sets from Π_n^1 under continuous functions. Additionally, Δ_n^1 is the intersection of Σ_n^1 and Π_n^1 . When the space X is not clear from the context, we add it in parentheses, e.g. $\Sigma_n^1(X)$.

We are mostly interested in the projective hierarchy for the space 2^ω with the topology of the Cantor set. This topology is induced by a metric, where the distance between two infinite bit strings is the inverse of the first position where they differ. We write $\Sigma_n^1(2^\omega)$ for the subsets of 2^ω that are in level Σ_n^1 of the projective hierarchy under this topology.

The main result. The main result of this paper is Theorem 1.1 from the introduction, which says that the MSO+U theory of 2^* is undecidable. The proof of Theorem 1.1 is by a reduction from the undecidability of MSO on 2^ω . Our proof uses a stronger undecidability version of MSO on 2^ω , where instead of full MSO we have a logic called *projective MSO*, where quantification is restricted to projective sets, as defined later in Section 2.1. We are unable to prove the projective MSO theory of 2^ω to be undecidable without any conditions, or even conditionally on the Continuum Hypothesis, but only assuming the stronger assumption that there exists a projective well-ordering of 2^ω of type ω_1 .

This assumption can be seen as a conjunction of two assumptions: the Continuum Hypothesis (the type ω_1 part) and that the well-ordering is “definable” in some sense (the projective part). As shown in [GS82] the MSO theory of 2^ω remains undecidable even without the Continuum Hypothesis. This does not help us, because our reduction to MSO+U crucially depends on the definability.

Before proving the theorem, we observe the following corollary.

Corollary 2.1. *If ZFC is consistent, then there is no algorithm which decides the MSO+U theory of 2^* and has a proof of correctness in ZFC.*

Proof. [The following proof is in ZFC] If ZFC is consistent, then Gödel's constructible universe L is a model of ZFC, as shown by Gödel (for a modern treatment of this topic see Chapter 13 and specifically Theorem 13.6 in [Jec02]). In Gödel's constructible universe, there exists a well-ordering of 2^ω of type ω_1 that is in level Δ_2^1 of the projective hierarchy on $2^\omega \times 2^\omega$ ([Jec02, Theorem 25.26]). Therefore, if ZFC is consistent, then by Theorem 1.1 it has a model where the MSO+U theory of 2^* is undecidable. \square

2.1 Projective MSO on $2^{\leq\omega}$, and its reduction to MSO+U on 2^*

For $n \leq \omega$, define the syntax of MSO_n to be the same as the syntax of MSO, except that instead of one pair of set quantifiers $\exists X$ and $\forall X$, there is a pair of quantifiers $\exists_i X$ and $\forall_i X$ for every $i \leq n$. To evaluate a sentence of MSO_n over a structure, we need a sequence $\{\mathcal{X}_j\}_{j \leq i}$ of families of sets, called the *monadic domains*. The semantics are then the same as for MSO, except that the quantifiers \exists_j and \forall_j are interpreted to range over subsets of the universe that belong to \mathcal{X}_j . First-order quantification is as usual, it can quantify over arbitrary elements of the universe. We write $\text{MSO}[\mathcal{X}_1, \mathcal{X}_2, \dots]$ for the above logic with the monadic domains being fixed to $\mathcal{X}_1, \mathcal{X}_2, \dots$. Standard MSO for structures with a universe Ω is the same as $\text{MSO}[\mathbf{P}(\Omega)]$, i.e. there is one monadic domain for the powerset of the universe. If Ω is equipped with a topology, we define *projective* MSO over Ω to be

$$\text{MSO}[\Sigma_1^1(\Omega), \Sigma_2^1(\Omega), \dots]$$

The expressive power of projective MSO is incomparable with the expressive power of MSO. Although projective MSO cannot quantify over arbitrary subsets, it can express that a set is in, say, Σ_1^1 .

Example 2.2. In the structure $2^{\leq\omega}$, being a node is first-order definable: a node is an element of the universe that is a proper prefix of some other element. Since there are countably many nodes, every set of nodes is Borel, and therefore in $\Sigma_1^1(2^{\leq\omega})$. Therefore, in projective MSO on $2^{\leq\omega}$ one can quantify over arbitrary sets of nodes. It is easy to see that a subset of $2^{\leq\omega}$ is in $\Sigma_n^1(2^{\leq\omega})$ if and only if it is a union of a set of nodes and a set from $\Sigma_n^1(2^\omega)$. It follows that projective MSO on $2^{\leq\omega}$ has the same expressive power as the logic

$$\text{MSO}[\mathbf{P}(2^*), \Sigma_1^1(2^\omega), \Sigma_2^1(2^\omega), \dots].$$

Example 2.3. In projective MSO on $2^{\leq\omega}$, one can say that a set of branches is countable. This is by using notions of interval, closed set, and perfect. A set of branches is open if and only if for every element, it contains some open interval around that element. A *perfect* is a set of branches which is closed (i.e. its complement is open) and contains no isolated points. The notions of open interval, closed set, and perfect are first-order definable. By [Kec95, Theorem 29.1], a set of branches is countable if and only if it is in $\Sigma_1^1(2^\omega)$ and does not contain any perfect subset, which is a property definable in projective MSO.

The following lemma shows that the projective MSO theory of $2^{\leq\omega}$ can be reduced to the MSO+U theory of 2^* .

Lemma 2.4. *For every sentence of projective MSO on $2^{\leq\omega}$, one can compute an equivalently satisfiable sentence of MSO+U on 2^* .*

The proof uses Theorem 5.1 from [HS12] and the following lemma.

Lemma 2.5. *Suppose that $L_1, L_2, \dots \subseteq A^\omega$ are definable in MSO+U, and let*

$$\mathcal{X}_i \stackrel{\text{def}}{=} \{f^{-1}(L_i) \mid f: 2^\omega \rightarrow A^\omega \text{ is a continuous function}\}.$$

Then for every sentence of MSO[P(2^), $\mathcal{X}_1, \mathcal{X}_2, \dots$] on $2^{\leq\omega}$, one can compute an equivalently satisfiable sentence of MSO+U on 2^* .*

Proof. The proof of this lemma is based on the observation that, using quantification over sets of nodes, one can quantify over continuous functions $2^\omega \rightarrow A^\omega$.

Call a mapping $f: 2^* \rightarrow A \cup \{\epsilon\}$ *proper* if on every infinite path in 2^* , the labelling f contains infinitely many letters different than ϵ . If f is proper then define $\hat{f}: 2^\omega \rightarrow A^\omega$ to be the function that maps a branch to the concatenation of values under f of nodes on the branch. It is not difficult to see that a function $g: 2^\omega \rightarrow A^\omega$ is continuous if and only if there exists a proper f such that $g = \hat{f}$, see e.g. Proposition 2.6 in [Kec95]. Since a mapping $f: 2^* \rightarrow A \cup \{\epsilon\}$ can be encoded as a family of disjoint sets $\{X_a \subseteq 2^*\}_{a \in A}$, one can use quantification over sets of nodes to simulate quantification over continuous functions $g: 2^\omega \rightarrow A^\omega$.

The reduction in the statement of the lemma works as follows. First-order quantification over branches is replaced by (monadic second-order) quantification over paths, i.e. subsets of 2^* that are totally ordered and maximal for that property. For a formula $\exists X \in \mathcal{X}_i. \varphi$, we replace the quantifier by existential quantification over a family of disjoint subsets $\{X_a\}_{a \in A}$ which encode a continuous function. In the formula φ , we replace a subformula $x \in X$, where x is now encoded as a path, by a formula which says that the image of x , under the function encoded by $\{X_a\}_{a \in A}$, belongs to the language L_i . In order to verify if a given element belongs to the language L_i definable in MSO+U on infinite words, we can use a formula of MSO+U on infinite trees. \square

Proof (of Lemma 2.4). Theorem 5.1 of [HS12] shows that there is an alphabet A such that for every $i \geq 1$, there is a language $L_i \subseteq A^\omega$ which is definable in MSO+U on infinite words and hard for $\Sigma_i^1(2^\omega)$. It is easy to check (see Appendix A.2, Fact 1) that L_i is in fact complete for $\Sigma_i^1(2^\omega)$. Apply Lemma 2.5 to these languages. By their completeness, the classes $\mathcal{X}_1, \mathcal{X}_2, \dots$ in Lemma 2.5 are exactly the projective hierarchy on 2^ω , and therefore Lemma 2.4 follows thanks to the observation at the end of Example 2.2. \square

Before we move on, we present an example of a nontrivial property that can be expressed in the projective MSO on $2^{\leq\omega}$.

Example: projective determinacy. A Gale-Stewart game with winning condition $W \subseteq 2^\omega$ is the following two-player game. For ω rounds, the players propose bits in an alternating fashion, with the first player proposing a bit in even-numbered rounds, and the second player proposing a bit in odd-numbered rounds. At the end of such a play, an infinite sequence of bits is produced, and the first player wins if this sequence belongs to W , otherwise the second player wins. Such a game is called *determined* if either the first or the second player has a winning strategy, see [Kec95, Chapter 20] or [Jec02, Chapter 33] for a broader reference. Martin [Mar75] proved that the games are determined if W is a Borel set.

It is not difficult to see that for every $i > 0$, the statement

“every Gale-Stewart game with a winning condition in Σ_i^1 is determined” (1)

can be formalised as a sentence φ_{det}^i of projective MSO on $2^{\leq\omega}$, see Appendix A.3. As we show below, the ability to formalise determinacy of Gale-Stewart games with winning conditions in Σ_1^1 already indicates that it is unlikely that projective MSO on $2^{\leq\omega}$ is decidable.

Indeed, suppose that there is an algorithm P deciding the projective MSO theory of $2^{\leq\omega}$ with a correctness proof in ZFC. Note that by Lemma 2.4, this would be the case if there was an algorithm deciding the MSO+U theory of 2^* with a correctness proof in ZFC. Run the algorithm on φ_{det}^1 obtaining an answer, either “yes” or “no”. The algorithm together with its proof of correctness and the run on φ_{det}^1 form a proof in ZFC resolving Statement (1) for $i = 1$. The determinacy of all Σ_1^1 games cannot be proved in ZFC, because it does not hold if $V=L$, see [Jec02, Corollary 25.37 and Section 33.9], and therefore P must answer “no” given input φ_{det}^1 .

This means that a proof of correctness for P would imply a ZFC proof that Statement (1) is false for $i = 1$. Such a possibility is considered very unlikely by set theorists, see [FFMS00] for a discussion of plausible axioms extending the standard set of ZFC axioms. A similar example regarding MSO(\mathbb{R}) and the Continuum Hypothesis was provided in [She75].

3 Undecidability of projective MSO on 2^ω

In this section we show that projective MSO is undecidable already on 2^ω with the lexicographic order. From the discussion in Example 2.2 it follows that the projective MSO theory of 2^ω reduces to the projective MSO theory of $2^{\leq\omega}$. Therefore, the undecidability result for 2^ω is stronger than for $2^{\leq\omega}$, in particular it implies the undecidability result for MSO+U from Theorem 1.1.

Theorem 3.1. *Assume that there is a projective well-ordering of 2^ω of type ω_1 . Then the projective MSO theory of 2^ω is undecidable.*

The proof of Theorem 3.1 is a minor adaptation of Shelah’s proof [She75] that, assuming the Continuum Hypothesis, the MSO theory of 2^ω is undecidable. In fact, Shelah already observed that such an adaptation is possible, in the following

remark on p. 410: “Aside from countable sets, we can use only a set constructible from any well-ordering of the reals.” To make the paper self-contained, we include a proof of Theorem 3.1.

Proof strategy. We use the name $\forall^*\exists^*$ *sentence* for a sentence of first-order logic in the prenex normal form that has a $\forall^*\exists^*$ quantification pattern. The vocabulary of graphs is defined to be the vocabulary with one binary predicate $E(x, y)$. Finally, an equality-free formula is one that does not use equality. The proof is by a reduction from the following satisfiability problem:

- **Input.** An equality-free $\forall^*\exists^*$ sentence over the vocabulary of graphs.
- **Question.** Is the sentence true in some undirected simple graph?

The above problem is undecidable by Theorem 1 in Section 9 of [Gur80].

Reducing from the above problem is one of the main differences between our proof and Shelah’s proof, which uses a reduction from the first-order theory of arithmetic $(\mathbb{N}, +, *)$. The other main difference is that we introduce two definitions, which we call modal graphs and Shelah graphs, which are only implicit in Shelah’s proof. Our intention behind these definitions is to give the reader a better intuition of what exactly is being coded into the MSO theory of 2^ω .

3.1 Modal graphs

Instead of encoding undirected simple graphs in projective MSO, it will be more convenient to encode a less rigid structure, which we call a *modal graph*¹. A modal graph consists of

- a partially ordered set of *worlds* with a least element;
- for every world I a set of *local vertices*² V_I ;
- for every world I a set of *local edges* $E_I \subseteq V_I \times V_I$

subject to the monotonicity property that $V_I \subseteq V_J$ and $E_I \subseteq E_J$ holds for every worlds $I \leq J$. Furthermore, for every I the local edges E_I are a symmetric irreflexive relation, i.e. modal graphs are simple and undirected.

We use first-order logic to describe properties of modal graphs, with the semantics relation denoted by

$$\mathcal{G}, I, val \models \varphi, \tag{2}$$

where φ is a formula of first-order logic, \mathcal{G} is a modal graph, I is a world in the modal graph, and val is a valuation that maps the free variables of φ to the local

¹ Another take on modality is presented in [GS82] using the language of forcing.

² We will only construct graphs where every world has the same local vertices, but we give the more general definition to match Kripke models for intuitionistic logic.

vertices V_I of the world I . The definition is by induction on the formula:

$\mathcal{G}, I, val \models E(x, y)$	iff	$(val(x), val(y)) \in E_I$
$\mathcal{G}, I, val \models \varphi \wedge \psi$	iff	$\mathcal{G}, I, val \models \varphi$ and $\mathcal{G}, I, val \models \psi$
$\mathcal{G}, I, val \models \varphi \vee \psi$	iff	$\mathcal{G}, I, val \models \varphi$ or $\mathcal{G}, I, val \models \psi$
$\mathcal{G}, I, val \models \neg\varphi$	iff	$\mathcal{G}, J, val \not\models \varphi$ for every $J \geq I$
$\mathcal{G}, I, val \models \exists x \varphi$	iff	$\mathcal{G}, J, val[x \rightarrow v] \models \varphi$ for some $J \geq I$ and $v \in V_J$
$\mathcal{G}, I, val \models \forall x \varphi$	iff	$\mathcal{G}, J, val[x \rightarrow v] \models \varphi$ for every $J \geq I$ and $v \in V_J$

The definition above is almost the same as Kripke's semantics for intuitionistic logic [Kri65]. The only difference is in the \exists quantifier: Kripke requires the world J to be equal to I . We say that a sentence (i.e. a formula without free variables) is satisfied in a modal graph if (2) holds with I being the least world and val being the empty valuation.

Example 3.2. A modal graph with one world is the same thing as an undirected simple graph. In this case, the standard semantics of first-order logic coincide with the semantics on modal graphs.

Example 3.3. Modal graphs satisfy more sentences of first-order logic than undirected simple graphs. In particular, if two existentially quantified sentences are satisfied in (possibly different) modal graphs, then their conjunction is also satisfied in the modal graph obtained by joining the two modal graphs by a common least world where there are no local edges.

The following lemma shows that for $\forall^*\exists^*$ -sentences, the answers are the same for the satisfiability problem in modal graphs and the satisfiability problem in simple undirected graphs. The same lemma would hold for directed graphs, and also for vocabularies with more predicates.

Lemma 3.4. *For every $\forall^*\exists^*$ sentence η over the vocabulary of graphs, η is satisfied in some undirected simple graph if and only if it is satisfied in some modal graph.*

Proof. The left-to-right implication is true for all sentences, not just $\forall^*\exists^*$ sentences, and follows from Example 3.2.

For the right-to-left implication, consider a $\forall^*\exists^*$ sentence

$$\eta = \forall x_1, \dots, x_k. \exists x_{k+1}, \dots, x_n. \alpha$$

where α is quantifier-free. For directed graphs G and H , we say that H is an η -extension of G if G is an induced subgraph of H , and for every valuation of the universally quantified variables of η that uses only vertices of G , there is a valuation of the existentially quantified variables of η which makes the formula α true, but possibly uses vertices from H .

Suppose that \mathcal{G} is a modal graph. For a world I and a subset V of the local vertices V_I , define $G_{I,V}$ to be the undirected simple graph where the vertices are V and the edges are local edges E_I restricted to $V \times V$. By monotonicity

of local edges, the set of edges in $G_{I,V}$ grows or stays equal as I grows. We say that $G_{I,V}$ is *stable* if $G_{I,V} = G_{J,V}$ holds for every $J \geq I$. The key properties of being stable are:

1. If $G_{I,V}$ is stable then for every valuation $val : \{x_1, \dots, x_n\} \rightarrow V$,

$$\mathcal{G}, I, val \models \alpha \quad \text{iff} \quad G_{I,V}, val \models \alpha.$$

In the equivalence above, the left side talks about semantics in modal graphs and the right side talks about semantics in simple undirected graphs.

2. For every world I and finite $V \subseteq V_I$, there exists a world $J \geq I$ such that $G_{I,V}$ is stable;
3. If $I \leq J$ are worlds and $V \subseteq W$ are such that $G_{I,V}$ and $G_{J,W}$ are stable, then $G_{I,V}$ is an induced subgraph of $G_{J,W}$.

Suppose that η is satisfied in \mathcal{G} .

Claim. There exists a sequence of worlds $I_1 \leq I_2 \leq \dots$ and a sequence $V_1 \subseteq V_2 \subseteq \dots$ of finite sets of vertices such that G_{I_i, V_i} is stable and η -extended by $G_{I_{i+1}, V_{i+1}}$ for every i .

This claim proves the lemma, since the limit, i.e. union, of the graphs G_{I_i, V_i} is a simple undirected graph that satisfies η .

Proof (of the claim). The sequence is constructed by induction; we only show the induction step. Suppose that I_i and V_i have already been defined. Let Γ_i be the finite set of valuations from the universally quantified variables x_1, \dots, x_k to the vertices V_i . Repeatedly using the assumption that \mathcal{G} satisfies η for every valuation in Γ_i , one shows that there exists a world $J \geq I$ such that every valuation $val \in \Gamma_i$ extends to a valuation

$$val' : \{x_1, \dots, x_n\} \rightarrow V_J \quad \text{such that} \quad \mathcal{G}, J, val' \models \alpha.$$

Define $V_{i+1} \subseteq V_J$ to be the finite set of vertices that are used by valuations of the form val' with val ranging over elements of Γ_i . Define $I_{i+1} \geq I_i$ to be the world, which exists by property 2 of stability, such that $G_{I_{i+1}, V_{i+1}}$ is stable. For quantifier-free formulas, the semantics in modal graphs are preserved when going into bigger worlds, and therefore

$$\mathcal{G}, I_{i+1}, val' \models \alpha$$

holds for every $val \in \Gamma_i$. By property 1 of stability, it follows that

$$G_{I_{i+1}, V_{i+1}}, val' \models \alpha.$$

Together with property 3 of stability, this implies that G_{I_i, V_i} is η -extended by $G_{I_{i+1}, V_{i+1}}$. □

□

3.2 Coding a modal graph in 2^ω

In this section, we describe how a modal graph can be coded in 2^ω . We use the name *interval* for a subset of 2^ω which consists of all branches that are lexicographically between some two distinct branches. Intervals defined this way are homeomorphic with 2^ω . Intervals are denoted I, J, K .

Define a *Shelah graph* to be two families \mathcal{V}, \mathcal{E} of subsets of 2^ω such that every set in \mathcal{V} is dense. For a Shelah graph, define its *associated modal graph* as follows. The worlds are the intervals in 2^ω , ordered by the opposite of inclusion, in particular the least world is the whole space 2^ω . The local vertices do not depend on the worlds: for every interval I , the local vertices V_I are \mathcal{V} (in particular a vertex is a subset of 2^ω). For an interval I and $V, W \in \mathcal{V}$, the local edge set E_I contains (V, W) if and only if

$$I \cap V \cap W = \emptyset \tag{3}$$

$$I \cap (V \cup W) = I \cap E \quad \text{for some } E \in \mathcal{E}. \tag{4}$$

It is easy to see that $E_I \subseteq E_J$ when interval J is included in interval I . Since worlds are ordered by the opposite of inclusion, this means that $I \leq J$ implies $E_I \subseteq E_J$. Every local edge set is symmetric because it is defined in terms of union and intersection. Every local edge is irreflexive because (3) implies $V \neq W$ (here we use density, since the dense sets V, W must have nonempty intersections with I). In other words the associated modal graph is a modal graph.

For a sentence φ of MSO_2 , and families \mathcal{V}, \mathcal{E} of subsets in 2^ω , we write

$$2^\omega, \mathcal{V}, \mathcal{E} \models \varphi$$

if φ holds, with the quantifiers $\exists_1 X$ and $\forall_1 X$ interpreted to range over sets in \mathcal{V} , and the quantifiers $\exists_2 X$ and $\forall_2 X$ interpreted to range over sets in \mathcal{E} . By using logic to formalise the definition of a Shelah graph, its associated modal graph, and the semantics of first-order logic on modal graphs, we get the following lemma.

Lemma 3.5. *For every sentence η of first-order logic over the vocabulary of graphs, one can compute a sentence $\hat{\eta}$ of MSO_2 such that*

$$2^\omega, \mathcal{V}, \mathcal{E} \models \hat{\eta}$$

if and only if $(\mathcal{V}, \mathcal{E})$ is a Shelah graph whose associated modal graph satisfies η .

The general idea in the undecidability result is to use $\hat{\eta}$ from the above lemma. The main problem is that a projective MSO sentence cannot begin saying “there exists a Shelah graph”, because a Shelah graph is described by an infinite (even uncountable) family of subsets of 2^ω . The solution to this problem, and the technical heart of the undecidability proof, is Proposition 3.6 below, which shows how to describe the infinite families $(\mathcal{V}, \mathcal{E})$ by using just four sets. The corresponding part in Shelah’s paper [She75] consists of Lemmas 7.6–7.9.

Proposition 3.6. *Assume that there exists a well-ordering of 2^ω of type ω_1 which belongs to $\Delta_k^1(2^\omega \times 2^\omega)$ for some k .*

Then there is a formula $\varphi_{\text{elem}}(V, Q, S)$ of projective MSO on 2^ω with the following property. If G is a countable undirected simple graph, then there are sets

$$Q_V, Q_E, S_V, S_E \subseteq 2^\omega, \quad (5)$$

such that the families

$$\mathcal{V} = \{V \subseteq 2^\omega : \varphi_{\text{elem}}(V, Q_V, S_V)\}, \quad \mathcal{E} = \{E \subseteq 2^\omega : \varphi_{\text{elem}}(E, Q_E, S_E)\} \quad (6)$$

are a Shelah graph whose associated modal graph satisfies the same equality-free $\forall^ \exists^*$ sentences as G .*

Furthermore, the formula φ_{elem} quantifies only over Σ_1^1 sets; the sets from (5) are in Σ_{k+4}^1 , and the families from (6) contain only countable sets.

We now use the proposition and the previous results to show the undecidability of projective MSO from Theorem 3.1.

Corollary 3.7. *Assume that there exists a projective well-ordering of 2^ω of type ω_1 . Let η be an equality-free $\forall^* \exists^*$ sentence over the vocabulary of graphs. Then the following conditions are equivalent:*

1. η is true in some undirected simple graph, with standard semantics of logic.
2. There are sets as in (5) such that the families \mathcal{V}, \mathcal{E} from (6) satisfy

$$2^\omega, \mathcal{V}, \mathcal{E} \models \hat{\eta}$$

where $\hat{\eta}$ is the sentence defined in Lemma 3.5.

3. η is true in some modal graph, with semantics of logic on modal graphs.

Proof. By the Löwenheim-Skolem theorem, if η is true in some undirected simple graph, then it is true in some countable undirected simple graph. Therefore, the implication $1 \Rightarrow 2$ follows from Proposition 3.6 and Lemma 3.5.

The implication $2 \Rightarrow 3$ follows from Lemma 3.5, which implies that η is true in some modal graph, namely the modal graph associated to the Shelah graph given by formula (6). The implication $3 \Rightarrow 1$ is the right-to-left implication in Lemma 3.4. \square

Proof (of Theorem 3.1). Condition 2 in the above corollary can be formalised by the formula of projective MSO on 2^ω

$$\exists S_V, Q_V, S_E, Q_E \in \Sigma_{k+4}^1. \tilde{\eta}$$

where k is the natural number from Proposition 3.6 and $\tilde{\eta}$ is the same as $\hat{\eta}$, except that instead of quantifying over a set $V \in \mathcal{V}$, it quantifies over a countable set V satisfying $\varphi_{\text{elem}}(V, Q_V, S_V)$; likewise for quantifying over $E \in \mathcal{E}$.

We have thus shown a reduction from the undecidable satisfiability problem for equality-free $\forall^* \exists^*$ sentences over undirected simple graphs to the theory of projective MSO on 2^ω . Therefore, the latter is undecidable. \square

4 Conclusions

We have shown that the $\text{MSO}+\text{U}$ theory of 2^* is undecidable, conditional on the existence of a projective well-ordering of 2^ω of type ω_1 . Apart from the obvious question about unconditional undecidability, a natural question is about the decidability of $\text{MSO}+\text{U}$ on infinite words: is the $\text{MSO}+\text{U}$ theory of the natural numbers with successor decidable? The methods used in this paper are strongly reliant on trees, so an undecidability proof would need new ideas to be adapted to the word case. Evidence for undecidability is that the topological hardness of $\text{MSO}+\text{U}$ on words is shown in [HS12] by encoding trees in words.

An interesting related problem [She75, Conjecture 7a] is the decidability of $\text{MSO}[\text{Borel}]$ on $2^{\leq\omega}$, i.e. the logic defined analogously to projective MSO except, that set quantification is over Borel sets only.

References

- [BC06] Mikołaj Bojańczyk and Thomas Colcombet. Bounds in ω -regularity. In *LICS*, pages 285–296, 2006.
- [Boj04] Mikołaj Bojańczyk. A bounding quantifier. In *CSL*, pages 41–55, 2004.
- [Boj11] Mikołaj Bojańczyk. Weak MSO with the unbounding quantifier. *Theory Comput. Syst.*, 48(3):554–576, 2011.
- [BT12] Mikołaj Bojańczyk and Szymon Toruńczyk. Weak $\text{MSO}+\text{U}$ over infinite trees. In *STACS*, pages 648–660, 2012.
- [CL08] Thomas Colcombet and Christof Löding. The non-deterministic Mostowski hierarchy and distance-parity automata. In *ICALP (2)*, pages 398–409, 2008.
- [Col09] Thomas Colcombet. The theory of stabilisation monoids and regular cost functions. In *ICALP (2)*, pages 139–150, 2009.
- [FFMS00] Solomon Feferman, Harvey M. Friedman, Penelope Maddy, and John R. Steel. Does mathematics need new axioms? *The Bulletin of Symbolic Logic*, 6(4):401–446, 2000.
- [GS82] Yuri Gurevich and Saharon Shelah. Monadic theory of order and topology in ZFC. *Annals of Mathematical Logic*, 23(2-3):179–198, 1982.
- [Gur80] Yuri Gurevich. Existential interpretation II. *Archiv für mathematische Logik und Grundlagenforschung*, 22(3-4):103–120, 1980.
- [HS12] Szczepan Hummel and Michał Skrzypczak. The topological complexity of $\text{MSO}+\text{U}$ and related automata models. *Fundamenta Informaticae*, 119(1):87–111, 2012.
- [HST10] Szczepan Hummel, Michał Skrzypczak, and Szymon Toruńczyk. On the topological complexity of $\text{MSO}+\text{U}$ and related automata models. In *MFCS*, pages 429–440, 2010.
- [Jec02] Thomas Jech. *Set Theory*. Springer-Verlag, 2002.
- [Kec95] Alexander Kechris. *Classical descriptive set theory*. Springer-Verlag, New York, 1995.
- [Kri65] S. A. Kripke. Semantical analysis of intuitionistic logic I. *Studies in Logic and the Foundations of Mathematics*, 1965.
- [Mar75] Donald A. Martin. Borel determinacy. *Annals of Mathematics*, 102(2):363–371, 1975.
- [Rab69] Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. of the American Math. Soc.*, 141:1–35, 1969.
- [She75] Saharon Shelah. The monadic theory of order. *The Annals of Mathematics*, 102(3):379–419, 1975.