On the decidability of MSO+U on infinite trees

Mikołaj Bojańczyk[1], Tomasz Gogacz[2], Henryk Michalewski[1] 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

$$\mathsf{U}X.\ \varphi(X),$$

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 MSO+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 where further developed into the theory of cost functions initiated by Colcombet in [Col09]. Cost functions can be seen as a special case of MSO+U in the sense that decision problems regarding cost functions, such as limitedness or domination, can be easily encoded into satisfiability of MSO+U formulae. This encoding need not be helpful, since the unsolved problems for cost functions get encoded into unsolved problems from MSO+U.

The added expressive power of MSO+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 MSO+U on infinite words; the particular fragment of MSO+U used in this reduction is decidable by [BC06]. An example of an important unsolved problem that reduces to MSO+U is the decidability of the nondeterministic parity index problem [CL08].

So far, most research on MSO+U has focussed on the weak variant, call it WMSO+U, where only quantification over finite sets is allowed. Satisfiability is decidable for WMSO+U over infinite words [Boj11] and infinite trees [BT12]. In a parallel submission to this conference, it is shown that WMSO+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 MSO+U can be too expressive was given in [HS12], where it was shown that MSO+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 MSO+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, MSO+U is undecidable on 2^*.

2 MSO+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 $\mathbf{\Sigma}^1_0 = \mathbf{\Pi}^1_0$ is the class of Borel sets, $\mathbf{\Pi}^1_n$ is the class of complements of sets from $\mathbf{\Sigma}^1_n$, and $\mathbf{\Sigma}^1_{n+1}$ is the class of images of sets from $\mathbf{\Pi}^1_n$ under continuous functions. Additionally, $\mathbf{\Delta}^1_n$ is the intersection of $\mathbf{\Sigma}^1_n$ and $\mathbf{\Pi}^1_n$. When the space X is not clear from the context, we add it in parentheses, e.g. $\mathbf{\Sigma}^1_n(X)$

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 $\mathbf{\Sigma}^1_n(2^\omega)$ for the subsets of 2^ω that are in level $\mathbf{\Sigma}^1_n$ 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 $\mathbf{\Delta}^1_2$ 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. □

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 $\text{MSO}[\mathcal{X}_1, \mathcal{X}_2, \ldots]$ for the above logic with the monadic domains being fixed to $\mathcal{X}_1, \mathcal{X}_2, \ldots$. Standard MSO for structures with a universe Ω is the same as $\text{MSO}[\mathsf{P}(\Omega)]$, 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

$$\text{MSO}[\mathbf{\Sigma}^1_1(\Omega), \mathbf{\Sigma}^1_2(\Omega), \ldots]$$

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, $\mathbf{\Sigma}^1_1$.

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 $\mathbf{\Sigma}^1_1(2^{\leq\omega})$. 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 $\mathbf{\Sigma}^1_n(2^{\leq\omega})$ if and only if it is a union of a set of nodes and a set from $\mathbf{\Sigma}^1_n(2^\omega)$. It follows that projective MSO on $2^{\leq\omega}$ has the same expressive power as the logic

$$\text{MSO}[\mathsf{P}(2^*), \mathbf{\Sigma}^1_1(2^\omega), \mathbf{\Sigma}^1_2(2^\omega), \ldots].$$

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 $\mathbf{\Sigma}^1_1(2^\omega)$ 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* $L_1, L_2, \ldots \subseteq A^\omega$ *are definable in* MSO+U*, and let*

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

Then for every sentence of $\text{MSO}[\mathsf{P}(2^*), \mathcal{X}_1, \mathcal{X}_2, \ldots]$ *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 $2^\omega \to A^\omega$.

Call a mapping $f \colon 2^* \to A \cup \{\epsilon\}$ *proper* if on every infinite path in 2^*, the labelling f contains infinitely many letters different than ϵ. If f is proper then define $\hat{f} \colon 2^\omega \to A^\omega$ 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 $g \colon 2^\omega \to A^\omega$ 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 $f \colon 2^* \to A \cup \{\epsilon\}$ 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 $g \colon 2^\omega \to A^\omega$.

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. □

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 $\mathbf{\Sigma}^1_i(2^\omega)$. It is easy to check (see the full version) that L_i is in fact complete for $\mathbf{\Sigma}^1_i(2^\omega)$. Apply Lemma 2.5 to these languages. By their completeness, the classes $\mathcal{X}_1, \mathcal{X}_2, \ldots$ 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. □

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

$$\text{“every Gale-Stewart game with a winning condition in } \mathbf{\Sigma}^1_i \text{ is determined”} \quad (1)$$

can be formalised as a sentence φ^i_{det} of projective MSO on $2^{\leq\omega}$ (see the full version). As we show below, the ability to formalise determinacy of Gale-Stewart games with winning conditions in $\mathbf{\Sigma}^1_1$ 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^1_{det} obtaining an answer, either “yes” or “no”. The algorithm together with its proof of correctness and the run on φ^1_{det} form a proof in ZFC resolving Statement (1) for $i = 1$. The determinacy of all $\mathbf{\Sigma}^1_1$ 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^1_{det}.

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	$\mathcal{G}, J, val[x \to v] \models \varphi$ for some $J \geq I$ and $v \in V_J$
$\mathcal{G}, I, val \models \forall x\ \varphi$	iff	$\mathcal{G}, J, val[x \to v] \models \varphi$ 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 the 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 $val : \{x_1, \ldots, x_n\} \to V$,

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

   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 $I_1 \leq I_2 \leq \ldots$ and a sequence $V_1 \subseteq V_2 \subseteq \cdots$ 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 $x_1, \ldots, x_k$ 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 $J \geq I_i$ such that every valuation $val \in \Gamma_i$ extends to a valuation

$$val' : \{x_1, \ldots, x_n\} \to V_J \qquad \text{such that} \qquad \mathcal{G}, J, val' \models \alpha.$$

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 $G_{I_{i+1},I_{i+1}}$. □

□

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 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

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

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

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 $\mathbf{\Delta}^1_k(2^\omega \times 2^\omega)$ 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*

$$Q_V, Q_E, S_V, S_E \subseteq 2^\omega, \tag{5}$$

such that the families

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

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 $\mathbf{\Sigma}^1_1$ sets; the sets from (5) are in $\mathbf{\Sigma}^1_{k+4}$, 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. □

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

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

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. □

4 Conclusions

We have shown that the MSO+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 MSO+U on infinite words: is the MSO+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 MSO+U on words is shown in [HS12] by encoding trees in words.

An interesting related problem [She75, Conjecture 7a] is the decidability of MSO[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 MSO+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 MSO+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 MSO+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.

5 Remaining proofs of lemmas

5.1 Formal translation in Lemma 2.5

In this section we formally prove Lemma 2.5. Assume that $L_1, L_2, \ldots \subseteq A^\omega$ is a sequence of MSO+U-definable sets. For $i > 0$ define

$$[f]_i \stackrel{\text{def}}{=} \left\{x \in 2^\omega : \hat{f}(x) \in L_i\right\},$$
$$\text{reduces}(L_i) \stackrel{\text{def}}{=} \{L \subseteq 2^\omega : L \text{ reduces continously to } L_i\}.$$

Proposition 2.6 in [Kec95] implies that

$$\{[f]_i : f \text{ is proper}\} = \text{reduces}(L_i). \tag{7}$$

Our translation inputs a formula of MSO[reduces(L_1), reduces$(L_2), \ldots$] and outputs a formula of MSO+U on 2^*. It interprets:

- a branch $x \in 2^\omega$ by the path $B_x = \{v \prec x\} \subseteq 2^*$,
- a set $X_i \in \text{reduces}(L_i)$ by a labelling $f_X^i : 2^* \to A \cup \{\epsilon\}$ such that $[f_X^i]_i = X_i$,
- a condition $v \prec x$ by $v \in B_x$,
- a condition $x \in X_i$ by checking that the formula defining L_i is true on the labelling f_X^i on the nodes in B_x.

Equation (7) says that the quantifications over $X_i \in \text{reduces}(L_i)$ and over proper labellings f_X^i are equivalent.

5.2 Completeness of languages $\mathrm{L}(\varphi_i)$

The authors of [HS12] provide examples of languages $L(\varphi_i)$ that are definable in MSO+U on infinite words and $L(\varphi_i)$ is $\mathbf{\Sigma}_i^1$-hard. In this work it is convenient to note the following fact.

Fact 1 *The languages $L(\varphi_i)$ from [HS12] are in fact $\mathbf{\Sigma}_i^1$-complete. Additionally, one can encode all these languages over one fixed alphabet.*

Proof. The translation to a fixed binary alphabet $\{0, 1\}$ is standard, we just fix a binary encoding of all the letters in the original alphabet and concatenate the codes. Such an encoding does not influence the topological complexity of the languages.

We refer on the definitions of formulae φ_i, see page 13 of [HS12]. We assume inductively that $\mathrm{L}(\varphi_{i-1}) \in \mathbf{\Pi}_i^1$ and show that $\mathrm{L}(\varphi_i) \in \mathbf{\Sigma}_i^1$, so in particular $\mathrm{L}(\varphi_i) \in \mathbf{\Pi}_{i+1}^1$.

Clearly $\mathrm{L}(\varphi_0)$ is a Borel language, so $\mathrm{L}(\varphi_0) \in \mathbf{\Pi}_1^1$. Note that, as shown in [HST10, Proposition 2], Conditions 1 and 2 of the definition of φ_i are Borel. Therefore, an infinite word satisfies φ_i if there exists a set G satisfying Conditions 1 and 2 and such that the bodies of the i-blocks of G form an infinite word satisfying $\neg\varphi_{i-1}$. By the inductive assumption, all these three conditions are $\mathbf{\Sigma}_i^1$ conditions, so $\mathrm{L}(\varphi_i)$ is a projection of a $\mathbf{\Sigma}_i^1$ language and it is itself $\mathbf{\Sigma}_i^1$. □

5.3 Formal construction of formulae expressing determinacy

For the sake of completeness we show how to express determinacy of Gale-Stewart games in MSO+U.

Assume that a formula $\mathrm{even}(v)$ (resp. $\mathrm{odd}(v)$) expresses that a given node is at the even (resp. odd) depth in the complete binary tree 2^*. By $S_0(v)$ and $S_1(v)$ we denote the respective successors of v in the tree, i.e. $S_i(v) = vi$. By $\epsilon \in 2^*$ we denote the root of the tree.

First, we define that a set of nodes encodes a strategy for the first player in the Gale-Stewart game:

$$\begin{aligned} S_{\mathrm{I}}(\sigma) = \epsilon \in \sigma \wedge & \\ & \forall w \in \sigma.\ \mathrm{even}(w) \ \Rightarrow (S_0(w) \in \sigma \Leftrightarrow S_1(w) \notin \sigma) \wedge \\ & \forall w \in \sigma.\ \mathrm{odd}(w) \ \Rightarrow (S_0(w) \in \sigma \wedge S_1(w) \in \sigma). \end{aligned}$$

The formula $S_{\mathrm{II}}(\sigma)$ defining a strategy for the second player is analogous except that the predicates even and odd are interchanged.

The following formula says that σ is a winning strategy for the first player for a winning condition $W \subseteq 2^\omega$:

$$\mathrm{win}_{\mathrm{I}}(\sigma, W) = S_{\mathrm{I}}(\sigma) \wedge \forall x \in 2^\omega.\ (\forall v \prec x.\ v \in \sigma) \Rightarrow x \in W.$$

Similarly we define

$$\mathrm{win}_{\mathrm{II}}(\sigma, W) = S_{\mathrm{II}}(\sigma) \wedge \forall x \in 2^\omega.\ (\forall v \prec x.\ v \in \sigma) \Rightarrow x \notin W.$$

Finally, Statement (1), namely the determinacy of all the Gale-Stewart games with winning conditions in $\mathbf{\Sigma}^1_i$ is expressed by

$$\forall W \in \mathbf{\Sigma}^1_i.\ \exists \sigma \in \mathsf{P}(2^*).\ \mathrm{win}_{\mathrm{I}}(\sigma, W) \ \vee \ \mathrm{win}_{\mathrm{II}}(\sigma, W).$$

By Lemma 2.4, this property can be effectively reduced to a certain sentence φ^i_{det} of MSO+U logic on the complete binary tree.

6 Qualifiers

The rest of the appendix is devoted to proving Proposition 3.6. The crucial concept behind Proposition 3.6 is an idea of a *qualifier* — a set that encodes an infinite family of subsets of 2^ω. This section is devoted to introducing the notion of a qualifier and proving its existence.

Unique sets, anti-unique sets, and qualifiers. Suppose that $\mathcal{D}$ is a family of dense countable subsets of 2^ω. Define the downward closure of $\mathcal{D}$ to be

$$\mathcal{D}{\downarrow} \stackrel{\mathrm{def}}{=} \{D : D \subseteq D' \text{ for some } D' \in \mathcal{D}\}.$$

– A set X is called *$\mathcal{D}$-unique* if there is some set $D \in \mathcal{D}\downarrow$ which contains the intersection of X with the union $\bigcup \mathcal{D}$.

$$\models \left(X \cap \bigcup \mathcal{D}\right) \in \mathcal{D}\downarrow .$$

– A set X is called *$\mathcal{D}$-anti-unique* if in every interval, X is either empty or not $\mathcal{D}$-unique.

$$\models \mathsf{G}\big(X = \emptyset \ \vee \ (X \cap \bigcup \mathcal{D}) \notin \mathcal{D}\downarrow \big)$$

We will apply the above definitions to X which are dense sets or perfect sets. When X is dense, the notion of $\mathcal{D}$-anti-unique simplifies to "in every interval, X is not $\mathcal{D}$-unique".

Definition 6.1. *Let $\mathcal{D}$ be a family of dense countable subsets of 2^ω. A set $Q \subseteq 2^\omega$ is called a* qualifier *for $\mathcal{D}$ if it is disjoint with the union of $\mathcal{D}$ and*

1. *every $\mathcal{D}$-unique perfect has countable intersection with Q,*
2. *every $\mathcal{D}$-anti-unique perfect has nonempty intersection with Q.*

Observe first that if Q is a qualifier for $\mathcal{D}$, then not only does Q intersect every $\mathcal{D}$-anti-unique perfect, but it is actually dense in it. Indeed, suppose that P is a $\mathcal{D}$-anti-unique perfect. Then after restricting to every interval which contains elements of P, P is also $\mathcal{D}$-anti-unique, and therefore in that interval P must intersect Q.

6.1 Existence of a qualifier

The goal of this section is Lemma 6.2, which says that for every countable dense set D and every countable family of dense subsets of D, there is a projective qualifier. Lemma 7.4 is a counterpart of this lemma in [She75]. In the following lemmas we will use a basic topological notion of a nowhere dense subset of 2^ω, that is a subset $C \subset 2^\omega$ such that the closure of C does not contain any interval. The Baire theorem guarantees in particular that in a compact space F a union of countably many nowhere dense subsets of F is not equal to F.

Lemma 6.2. *If there exists a well-ordering $\preceq$ of 2^ω of type ω_1 such that $\preceq \in \boldsymbol{\Delta}^1_k(2^\omega \times 2^\omega)$ then every countable family $\mathcal{D}$ of dense sets such that $\bigcup \mathcal{D}$ is countable admits a $\boldsymbol{\Sigma}^1_{k+4}$ qualifier.*

Proof. The following proof is rather routine from the point of view of the descriptive set theory and is an illustration of a general claim that every set constructed using a formula of set theory which refers only to Borel and projective sets is a member of certain level of the projective hierarchy. The only unusual aspect of the proof is that we use a peculiar projective set, namely a well-ordering of 2^ω of type ω_1. In the proof we will need the following

Claim. For every $\mathcal{D}$-anti-unique perfect P and $\mathcal{D}$-unique perfect P', the intersection $P \cap P'$ is nowhere dense in P.

Proof. (of the Claim) Suppose towards a contradiction, that $P' \cap I$ is dense in $P \cap I$ for some interval I. Since P' is a closed set, the intersection $P' \cap I$ would contain $P \cap I$. Therefore, the $\mathcal{D}$-unique perfect P' would contain the $\mathcal{D}$-anti-unique perfect $P \cap I$, a contradiction. $\square$

Let $U_{\text{closed}} \subseteq 2^\omega \times 2^\omega$ be a universal set for closed subsets of 2^ω (see [Kec95, Theorem 22.3]). This means, that U_{closed} is closed in $2^\omega \times 2^\omega$ and for every closed subset $F \subseteq 2^\omega$ there exists $x \in 2^\omega$ such that $F = U_x$ where $U_x = \{y \in 2^\omega : (x, y) \in U_{\text{closed}}\}$. We define

$$\begin{aligned} P_{\text{perfect}} &= \{x \in 2^\omega : \ U_x \text{ is perfect}\}, \\ P_{\text{unique}} &= \{x \in P_{\text{perfect}} : \ U_x \text{ is } \mathcal{D}\text{-unique}\}, \\ P_{\text{anti-unique}} &= \{x \in P_{\text{perfect}} : \ U_x \text{ is } \mathcal{D}\text{-anti-unique}\}. \end{aligned}$$

So far all the definitions work without any special set-theoretic assumptions and at the same time through unfolding of the formulae one can routinely check that all involved sets are projective. Now we use the assumption that 2^ω admits a $\boldsymbol{\Delta}^1_k$ well-ordering of type ω_1. This means that there exists a binary relation $\preceq \subseteq 2^\omega \times 2^\omega$ such that $\{y \in 2^\omega : y \preceq x\}$ is countable for every $x \in 2^\omega$. Define

$$R = \left\{(x, y) \in 2^\omega \times 2^\omega : \ y \in \bigcup \left\{U_z : z \preceq x \wedge z \in P_{\text{unique}}\right\}\right\}.$$

For a given point $x \in 2^\omega$ the set R_x is a union of all $\mathcal{D}$-unique perfects indexed by z preceding x in the well-ordering $\preceq$. In order to satisfy the definition of a $\mathcal{D}$-qualifier, if x is in $P_{\text{anti-unique}}$, that is U_x is a $\mathcal{D}$-anti-unique perfect, we want to include into the qualifier at least one element $y \in U_x \setminus R_x \setminus \bigcup \mathcal{D}$.

Fact 2 *The set $U_x \setminus R_x \setminus \bigcup \mathcal{D}$ is nonempty for $x \in P_{\text{anti-unique}}$.*

Proof. There are countably many perfects U_z with $z \preceq x$. The intersections $U_z \cap U_x$ are nowhere dense in U_x by the above Claim. Therefore, by the Baire theorem, it follows that U_x cannot be exhausted by the union of all the intersections $U_x \cap U_z$, ranging over $\mathcal{D}$-unique perfects U_z for $z \preceq x$. Subtracting the countable set $\bigcup \mathcal{D}$ will also not exhaust U_x. So there must be some element left. $\square$

Since we have an access to a projective well-ordering of 2^ω, we can take the minimal $y \in U_x \setminus R_x \setminus \bigcup \mathcal{D}$ and still remain within the projective hierarchy. More precisely, we define the qualifier for $\mathcal{D}$ as

$$Q = \left\{y \in 2^\omega : \exists_{x \in P_{\text{anti-unique}}} \ y = \min_{\preceq} \ U_x \setminus R_x \setminus \bigcup \mathcal{D}\right\}.$$

A precise computation shows that we can obtain $Q \in \boldsymbol{\Sigma}^1_{k+4}$. $\square$

7 Using qualifiers to describe a Shelah graph

This section is devoted to proving Proposition 3.6. Suppose that $\mathcal{V}_0$ and $\mathcal{E}_0$ are countable families of sets such that

- the union $\bigcup \mathcal{V}_0$ is countable;
- the sets in $\mathcal{V}_0$ are pairwise disjoint and dense;
- every set in $\mathcal{E}_0$ is the union of some two distinct sets in $\mathcal{V}_0$.

Note that not every union of two distinct sets in $\mathcal{V}_0$ needs to belong to $\mathcal{E}_0$. Define the *Shelah graph generated by* $\mathcal{V}_0$ *and* $\mathcal{E}_0$ as follows. The families $\mathcal{V}$ (resp. $\mathcal{E}$) contain those subsets of $\bigcup \mathcal{V}_0$ that are globally finally equal to some element of $\mathcal{V}_0$ (resp. $\mathcal{E}_0$):

$$\mathcal{V} = \left\{ V \subseteq \bigcup \mathcal{V}_0 : \models \mathsf{GF}\ V \in \mathcal{V}_0 \right\}, \tag{8}$$

$$\mathcal{E} = \left\{ E \subseteq \bigcup \mathcal{E}_0 : \models \mathsf{GF}\ E \in \mathcal{E}_0 \right\}. \tag{9}$$

Fact 3 *The pair* $(\mathcal{V}, \mathcal{E})$ *is a Shelah graph.*

Proof. It is enough to verify the axioms of Shelah graphs from Subsection 3.2. Since all the axioms start with the GF prefix so we can always satisfy them by restricting to an interval where the given sets are equal to some elements of $\mathcal{V}_0$. Then, the conditions follow directly from the properties of $\mathcal{V}_0$ and $\mathcal{E}_0$. □

Remark 7.1. One can observe that not every Shelah graph is generated by some $\mathcal{V}_0$ and $\mathcal{E}_0$ (hint: two different graphs can be encoded on the two canonical halves of 2^ω).

7.1 Defining elements of $\mathcal{D}\!\downarrow$

The following proposition shows how to quantify over elements of $\mathcal{D}\!\downarrow$ using a qualifier Q. It corresponds to Lemma 7.6 of [She75].

Proposition 7.2. *Suppose that* $(\mathcal{V}, \mathcal{E})$ *is a Shelah graph generated by* $\mathcal{V}_0$ *and* $\mathcal{E}_0$*. Let* $\mathcal{D}$ *be either* $\mathcal{V}_0$ *or* $\mathcal{E}_0$*, and let* Q *be a qualifier for* $\mathcal{D}$*. The following conditions are equivalent for every dense set* X *which is included in the union of* $\mathcal{D}$*.*

1. *Globally finally* X *is in the downward closure of* $\mathcal{D}$*:*

$$\models \mathsf{GF}\ X \in \mathcal{D}\!\downarrow.$$

2. *If* $\mathcal{A} = \{A_1, A_2, A_3\}$ *is a family of at most three disjoint subsets of* X *which are dense in some common interval, i.e.*

$$\models \mathsf{FG} \bigwedge_{A_i \in \mathcal{A}} A_i \neq \emptyset$$

then there is some perfect P *which is disjoint with* Q*, and such that each of the sets from* $\mathcal{A}$ *is dense in* P*.*

The main point of this proposition is the following

Corollary 7.3. *The second condition above is* MSO*-definable, in terms of: a set X, a qualifier Q, and the union of a family $\mathcal{D}$.*

The rest of this subsection is devoted to showing Proposition 7.2.

We first prove the top-down implication. Assume Condition 1 and take a family $\mathcal{A}$ as in the assumption of Condition 2, all of whose elements are dense in some interval I. By Condition 1, on some subinterval J of I the set X is included in some set from $\mathcal{D}$:

$$J \models X \subseteq D \qquad \text{for some interval } J \subseteq I \text{ and } D \in \mathcal{D}$$

In particular, in the interval J, all sets from $\mathcal{A}$ are included in D:

$$J \models A \subseteq D \qquad \text{for every } A \in \mathcal{A}.$$

We will now need the following combinatorial lemma.

Lemma 7.4. *Let P be a perfect. Let $Z \subseteq 2^\omega$ be countable, and $\mathcal{A}$ be a nonempty countable family of subsets of P that are dense in P but disjoint from Z. There is a perfect included in P which is disjoint from Z, and such that every set from $\mathcal{A}$ is dense in P.*

Proof. This useful lemma appears in [She75] as a part of the proof of Lemma 7.6.

We construct inductively a countable set P_∞ of elements in P in such a way, that the perfect P' defined as the closure of P_∞ fulfils the requirements of the lemma. There are the following three aspects of the construction which we must control:

1. Some elements of P may be already in Z — we must avoid them and make sure, that limits of points in P_∞ also avoid Z; once this is achieved, P' will be disjoint with Z. We achieve this effect making sure that for a given $z \in Z$ there is a distance $dist_z > 0$ such that if a new element x is added to P_∞ then the distance $\mathrm{d}(z, x) \geq dist_z$. This means, that if x is a limit of a sequence of elements in P_∞, then the distance of z and x is not smaller than $dist_z$. Hence, the limit cannot be in Z.
2. The set P' must be perfect, that is for every element $x \in P_\infty$ and for every prefix $p \prec x$ we must add two different elements of P_∞ extending p; in order to achieve this effect we will pick the extensions from sets of the form $A \cap P$ for $A \in \mathcal{A}$, since we know upfront that they are dense in P and disjoint with Z.
3. Every set $A \in \mathcal{A}$ must be dense in P'; we achieve this effect returning to each set $A \in \mathcal{A}$ infinitely many times and picking sufficiently many elements from A.

From this high-level overview let us move to some technical details. Choose an enumeration $z_0, z_1, \ldots$ of Z, and an enumeration $A_0, A_1, \ldots$ of $\mathcal{A}$ such that every set from $\mathcal{A}$ appears infinitely often in the enumeration. For $n \in \mathbb{N}$ we define by induction a finite set of elements $P_n \subseteq A_n \subseteq P$; the above mentioned set P_∞ will be defined as $\bigcup_{n\in\mathbb{N}} P_n$. For $n = 0$ we define P_0 to be an arbitrary element $x \in A_0$. Let $dist_{z_0} = \mathrm{d}(x, z_0)$. Let $d_0 = 0$. As mentioned in the above Condition 1 we will make sure, that the new elements added in subsequent steps of the construction will be at the distance $\mathrm{d}(x, z_0)$ from z_0.

Suppose that the sets $P_0, \ldots, P_{n-1}$ have already been defined; let $P_{<n}$ denote their union. Suppose also, that for $k < n$ for every $x \in P_{<n}$ we have $\mathrm{d}(x, z_k) \geq dist_{z_k}$. Let $dist_{z_n} = \min\{\mathrm{d}(x, z_n) : x \in P_{<n}\}$. The value $dist_{z_n}$ is bigger than 0 since the set $P_{<n}$ is finite and Z is disjoint with each A_i. Let $d_n \in \mathbb{N}$ be the minimal depth such that the finitely many elements

$$P_{<n} \cup \{z_1, \ldots, z_n\}$$

of 2^ω can be distinguished on coordinates smaller than d_n; we also assume that $d_n \geq \max(d_{n-1}, n)$. By assumption on A_n being dense in P, for every $x \in P_{<n}$, we can find two distinct elements x_1, x_2 of A_n which agree with x on coordinates up to d_n. Put these two elements into P_n, ranging over all $x \in P_{<n}$. We notice, that for $k \leq n$ we have $\mathrm{d}(x_i, z_k) = \mathrm{d}(x, z_k) \geq dist_{z_k}$. This ends the inductive construction and shows that the invariant $dist_{z_k}$ was preserved.

Define P_∞ to be the union of all P_n. Since each P_n is included in P, and P is a closed set, it follows that the closure P' of P_∞ is included in P. We claim that P' also has other properties required by the lemma.

1. Invariant $dist_{z_n}$ guarantees that $z_n \notin P'$ ($n \in \mathbb{N}$), hence Z is disjoint with P'.
2. Let us notice that P_∞ is indeed perfect. Consider $x \in P_n \subset P_\infty$. Fix a prefix p of x. At stage $k = \max(\mathrm{length}(p), n+1)$ of the inductive construction we added two extensions of x restricted to d_k and since $d_k \geq k$, these two extensions will also extend the prefix p.
3. It remains to check that every set $A \in \mathcal{A}$ is indeed dense in P'. Consider any $x \in P_n \subset P_\infty$ and fix a prefix p of x. It is enough to check, that there exists $x' \in P_\infty \cap A$ extending p. Since in the enumeration $A_0, A_1, \ldots$ the set A appears infinitely many times, we fix k such that $k \geq \max(\mathrm{length}(p), n+1)$ and $A_k = A$. From the inductive construction follows, that in the k-th step we added two elements from A_k extending the prefix p.

□

No we apply Lemma 7.4 to the perfect being the interval J, the family $\mathcal{A}$, and

$$Z = \left(\bigcup \mathcal{D}\right) \setminus D.$$

The result is a perfect in the interval J, call it P_1, where each of the sets from $\mathcal{A}$ is dense, and which intersects the union of $\mathcal{D}$ only on the elements of D. The last property implies that P_1 is $\mathcal{D}$-unique, and therefore it has countable intersection with Q. Apply Lemma 7.4 once again, with the perfect being P_1, the family $\mathcal{A}$, and the set Z being the previous set Z plus $P_1 \cap Q$. The resulting perfect $P \subseteq P_1$ is now disjoint with Q, as required by Condition 2 of Proposition 7.2.

We now prove the bottom-up implication. Assuming that X satisfies Condition 2 of Proposition 7.2, we need to prove that

$$\models \mathsf{GF}\ X \in \mathcal{D}{\downarrow}\,. \tag{10}$$

For this we will use the following lemma.

Lemma 7.5. *Assuming Condition 2 of Proposition 7.2, every interval I contains a subinterval J such that, with at most two exceptions, every set $V \in \mathcal{V}_0$ is such that $V \cap X$ is nowhere dense in J, i.e.*

$$\models \mathsf{GF} \bigvee_{V_1,V_2\in\mathcal{V}_0} \ \bigwedge_{V\in\mathcal{V}_0\setminus\{V_1,V_2\}} \mathsf{GF}\ V \cap X = \emptyset$$

Proof. Towards a contradiction, suppose that for some interval I, in every subinterval $J \subseteq I$, there are three different sets $V_1, V_2, V_3 \in \mathcal{V}_0$ such that each of the sets $V_i \cap X$ is dense in some (possibly different depending on i) subinterval of J.

$$I \models \mathsf{G} \bigvee_{V_1\neq V_2\neq V_3\in\mathcal{V}_0} \ \bigwedge_{i\in\{1,2,3\}} \mathsf{FG}\ V_i \cap X \neq \emptyset$$

Claim. There exists a subinterval J of I and three different sets $V_1, V_2, V_3 \in \mathcal{V}_0$, such that each $V_i \cap X$ is dense in J.

Proof. First, there must be a subinterval $I_1 \subseteq I$ and a set $V_1 \in \mathcal{V}$ such that V_1 is dense in I_1. Since $I_1 \subseteq I$ there must be a subset $I_2 \subseteq I_1$ and a set $V_2 \in \mathcal{V}$ such that V_2 is dense in I_2. Finally since $I_2 \subseteq I$ there must be a subinterval $J \subseteq I_2$ and a set $\mathcal{V}_3$ such that $\mathcal{V}_3$ is dense in J. □

By Condition 2 of Proposition 7.2, all the sets $V_1 \cap X$, $V_2 \cap X$, $V_3 \cap X$ are dense in some perfect P disjoint with Q. By the assumption that all sets in $\mathcal{V}_0$ are disjoint, and that every set from $\mathcal{D}$ can intersect at most two sets from $\mathcal{V}_0$, the perfect P is $\mathcal{D}$-anti-unique. But that means that P intersects Q, contradicting the assumption. □

Now we are returning to the proof of Proposition 7.2 and will prove (10). Take any $I \subseteq 2^\omega$ and let $J \subseteq I$ be given by the above lemma. Choose some enumeration

$$V_1, V_2, V_3, V_4, \cdots$$

of $\mathcal{V}_0$ such that the at most two exceptions are among V_1, V_2. We want to show that $J \models X \in \mathcal{D}{\downarrow}$. Define Y to be the set of elements of X that do not belong to the exceptions.

Claim. Y is not dense in J.

Proof. Towards a contradiction, suppose that Y is dense in J. The intersection of X with every set from $V_3, V_4, \dots$ is nowhere dense in J. This implies that subtracting any finite union $V_3 \cup \cdots \cup V_n$ from Y still yields a set that is dense in J. Therefore, we can inductively extract a subset of $Z \subseteq Y$ which is still dense in J, but which uses at most one element from each set in $\mathcal{V}_0$. In particular, Z is $\mathcal{D}$-anti-unique. This contradicts Lemma 7.4 applied to $\mathcal{A} = \{Z\}$, which says that Z is dense in some perfect disjoint with Q, therefore Z is dense in some perfect that is not $\mathcal{D}$-anti-unique, and therefore Z cannot be $\mathcal{D}$-anti-unique. □

To finish the proof of Proposition 7.2, we will find a subinterval of J where X is contained in some set from $\mathcal{D}$.

Since Y is not dense in the interval J, on some subinterval of J, the set X contains no elements from Y, and therefore contains only elements from $V_1 \cup V_2$. Since X is dense, there must be some nonempty $\mathcal{A} \subseteq \{V_1, V_2\}$ (three possibilities for this $\mathcal{A}$) such that on some subinterval K of J, X is included in the union of $\mathcal{A}$ and each set from $\mathcal{A}$ is dense in X. By Condition 2 of Proposition 7.2, there is a perfect P in the interval K disjoint with Q where all the sets from $\mathcal{A}$ are dense. Since the perfect is disjoint with Q, it is not $\mathcal{D}$-anti-unique, and therefore on some subinterval K' of K it is contained in some set from $\mathcal{D}$. Therefore, in this interval K', the union of $\mathcal{A}$, which contains X, is also contained in some set from $\mathcal{D}$.

7.2 Defining elements of $\mathcal{D}$

When reading the following lemma, think of $\mathcal{D}$ as being either $\mathcal{V}_0$ or $\mathcal{E}_0$, i.e. generators of a Shelah graph. Note that in the case $\mathcal{D} = \mathcal{V}_0$ every two distinct elements of $\mathcal{D}$ are disjoint; whereas in the case of $\mathcal{D} = \mathcal{E}_0$ every two distinct elements of $\mathcal{D}$ are incomparable with respect to inclusion. This lemma corresponds to Lemma 7.7 of [She75].

Lemma 7.6. *Let $\mathcal{D}$ be a countable family of dense countable sets such that in every interval all the sets from $\mathcal{D}$ are incomparable with respect to inclusion:*

$$\models \mathsf{G} \bigwedge_{X \neq Y \in \mathcal{D}} X \not\subseteq Y$$

Then the following conditions are equivalent for a set $X \subseteq 2^\omega$.

1. *Globally finally X belongs to $\mathcal{D}$:*

$$\models \mathsf{GF}\ X \in \mathcal{D}$$

2. (a) *Globally finally X belongs to the downward closure $\mathcal{D}{\downarrow}$*

$$\models \mathsf{GF}\ X \in \mathcal{D}{\downarrow},$$

i.e. X satisfies Condition 1 of Proposition 7.2.

(b) *If Condition 2a holds for some $Y \supseteq X$, then $Y \setminus X$ is nowhere dense.*

Proof. For the top-down implication, assume that Condition 1 holds. Condition 2a is immediate. For Condition 2b, suppose that $Y \supseteq X$ satisfies Condition 2a. We will show that globally finally X is equal to Y, i.e.

$$\models \mathsf{GF}\ X = Y,$$

and therefore $Y \setminus X$ is nowhere dense. Let I be an interval. By Condition 1, X must be equal to some set $D \in \mathcal{D}$ on some subinterval. Because Y satisfies Condition 2a, on some smaller subinterval it must also be equal to some $D' \in \mathcal{D}$, but then because the sets in $\mathcal{D}$ are incomparable with respect to inclusion in every subinterval, it follows that $D = D'$, and therefore Y agrees with X on the smaller subinterval.

Consider now the bottom-up implication. Let X be a set which satisfies Conditions 2a and 2b, and let I be an interval. We will show that on some subinterval of I, X is equal to a set from $\mathcal{D}$. By Condition 2a, on some subinterval $J \subseteq I$, the set X is included in some $D \in \mathcal{D}$. By Condition 2b, the difference $D \setminus X$ is nowhere dense on J, and therefore there must be a subinterval of J where D is equal to X. □

7.3 Proof of Proposition 3.6

We are now ready to prove Proposition 3.6, and thus finish the proof of Theorem 3.1.

As noted in Corollary 7.3, Condition 2 of Proposition 7.2 can be expressed by a formula of projective MSO, which takes as parameters a set X, a qualifier Q, and the union of the set $\mathcal{D}$. Call this formula $\psi(X, Q, S_D)$.

Let $\varphi_{\text{elem}}(X, Q, S_D)$ be the formula expressing Condition 2 of Lemma 7.6: it says that $\psi(X, Q, S_D)$ holds, and there is no set $Y \supseteq X$ which is included in S_D, such that $\psi(Y, Q, S_D)$ holds and $Y \setminus X$ is dense in some interval.

Choose a countable family $\mathcal{V}_0$ of disjoint dense countable sets, with one set V_v for each vertex v of the graph G. Define $\mathcal{E}_0$ to be the family of sets $V_v \cup V_w$ ranging over pairs of vertices v and w that are connected by an edge in the graph G. Let Q_V be a qualifier for $\mathcal{V}_0$ and let Q_E be a qualifier for $\mathcal{E}_0$, both of these exist by Lemma 6.2.

Define S_V to be the union of $\mathcal{V}_0$ and S_E to be the union of $\mathcal{E}_0$. This completes the definition of the sets in (5).

Let now $\mathcal{D}$ be either $\mathcal{V}_0$ or $\mathcal{E}_0$, and let Q be a qualifier for $\mathcal{D}$. By Proposition 7.2 and Lemma 7.6, $\varphi_{\text{elem}}(X, Q, \bigcup \mathcal{D})$ holds if and only if in every interval there is some subinterval where X is equal to an element of $\mathcal{D}$, i.e.

$$\varphi_{\text{elem}}\left(X, Q, \bigcup \mathcal{D}\right) \text{ holds} \quad \text{if and only if} \quad \models \mathsf{GF}\ X \in \mathcal{D}.$$

In other words, the families

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

as in (6) form the Shelah graph $(\mathcal{V}, \mathcal{E})$ generated by $\mathcal{V}_0$ and $\mathcal{E}_0$ (see (8) and (9)).

It is easy to check that $(\mathcal{V}, \mathcal{E})$ is in fact a Shelah graph. Axiom (**??**) (density) for sets in $\mathcal{V}$ follows directly from the definition. Lemma 7.6 and disjointness of the elements of $\mathcal{V}_0$ give Axiom (**??**). Once we know that globally finally every two elements of $\mathcal{V}$ are equal to some sets $V_1, V_2 \in \mathcal{V}_0$, the edge relation between them is settled down (equivalent to the statement $V_1 \cup V_2 \in \mathcal{E}_0$). Hence, Axiom (**??**) is also true.

Additionally, our construction ensures that the union of the family $\mathcal{V}$ is countable, in particular all the sets $V \in \mathcal{V}$ and $E \in \mathcal{E}$ are countable. Also, the formula φ_{elem} quantifies only over $\mathbf{\Sigma}^1_1$ sets. As stated in Lemma 6.2, the sets Q_V, Q_E, S_V, S_E are in $\mathbf{\Sigma}^1_{k+4}$ (in fact S_V, S_E are countable).

Let $\mathcal{G}$ be the modal graph associated to $(\mathcal{V}, \mathcal{E})$. What remains is to show the following lemma.

Lemma 7.7. *For every $\forall^*\exists^*$ equality-free sentence*

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

with α quantifier-free the following conditions are equivalent:

- *G satisfies η with the standard semantics of first-order logic,*
- *$\mathcal{G}$ satisfies η with the semantics of first-order logic on modal graphs.*

Suppose that G satisfies η. Let $I \subseteq 2^\omega$ be an arbitrary interval and let $W_1, \ldots, W_k \in \mathcal{V}$ be a valuation of the universally quantified variables such that $\mathcal{G}$ is total on $\{W_1, \ldots, W_k\}$ in I. We need to show that

$$\mathcal{G}, I, W_1, \ldots, W_k \models \exists x_{k+1}, \ldots, x_n.\ \alpha.$$

By Lemma 7.6 there is a subinterval $J \subseteq I$ such that $W_1, \ldots, W_k$ are equal to some elements $V_1, \ldots, V_k \in \mathcal{V}_0$ relatively to J. Those sets correspond to vertices $v_1, \ldots, v_k$ of G. Hence, there are witnesses $v_{k+1}, \ldots, v_n$ in G for existentially quantified variables. By the construction of $(\mathcal{V}, \mathcal{E})$ we know that $\mathcal{G}$ is total on the set $\{W_1, \ldots, W_k, V_{k+1}, \ldots, V_n\}$ in J. The edge relation on these elements given by $\mathcal{E}_0$ is the same as the edge relation in G, therefore

$$\mathcal{G}, J, W_1, \ldots, W_k, V_{k+1}, \ldots, V_n \models \alpha.$$

The proof of the second implication is almost the same. Suppose that $\mathcal{G}$ satisfies η. Let $v_1, \ldots, v_k \in G$ be a valuation of the universally quantified variables of η. Since

$$\mathcal{G}, 2^\omega, V_1, \ldots V_k \models \exists x_{k+1}, \ldots, x_n.\ \alpha,$$

there is an interval I and a valuation $W_{k+1}, \ldots, W_n$ of the existentially quantified variables such that

$$\mathcal{G}, I, V_1, \ldots, V_k, W_{k+1}, \ldots, W_n \models \alpha.$$

By Lemma 7.6, we can further restrict to a subinterval J where the sets $W_{k+1}, \ldots, W_n$ belong to $\mathcal{V}_0$. By monotonicity of modal graph semantics for quantifier-free formulas, we have

$$\mathcal{G}, J, V_1, \ldots, V_k, W_{k+1}, \ldots, W_n \models \alpha.$$

so they provide a valuation needed for graph G.