Automata for MSO over Infinite Trees with Quantification over Borel Sets of Branches

Mikołaj Bojańczyk[1,*], Antonio Casares[2,†], Sven Manthe[3,‡], and Paweł Parys[1,§]

[1] Institute of Informatics, University of Warsaw, Poland
{bojan,parys}@mimuw.edu.pl
[2] University of Kaiserslautern-Landau, Germany
antonio.casares@rptu.de
[3] University of Bonn, Germany
sven.manthe@uni-muenster.de

Abstract

Rabin's Tree Theorem says that the MSO theory of the infinite binary tree 2^* is decidable. Shelah showed that MSO logic becomes undecidable if this tree is extended to $2^{\leq\omega}$, that is, by allowing quantification over sets of infinite branches. A longstanding open problem is whether the decidability can be recovered in $2^{\leq\omega}$ by restricting set quantification to Borel sets. We make some progress in this direction, by identifying a suitable automaton model, and showing that most of the automata-theoretic approach to Rabin's Theorem can be extended to the new framework. The only missing part is a conjecture about finite-memory determinacy in certain games. This paper states and explores the conjecture. We prove it in some restricted cases, and give lower bounds on the memory required in those games.

This is a short abstract based on a paper accepted to LICS 2026.

One of the landmark results of logic in computer science is Rabin's Tree Theorem [9], which states the decidability of the MSO theory of the infinite binary tree, which is the set 2^* equipped with the prefix relation. This is a powerful theory—for instance, it interprets the MSO theory of the rational numbers with the usual order [9, Thm. 2.1], which is thus decidable—and it is close to the border of undecidability. On the other side of this border we find the Cantor set 2^ω, equipped with the lexicographic order, and also the order of the reals. Both of these structures have an undecidable MSO theory (as shown by Shelah [10, Theorem 7], assuming the Continuum Hypothesis; this assumption was later removed in [6], i.e., undecidability can be proved in ZFC only). The undecidability proofs use sets that are complicated from a topological point of view, and Shelah conjectured [10, Conjectures 7A and 7B] (see also [4, Open Question, p. 503][11, 2.22]) that decidability can be recovered by restricting set quantification to range over Borel sets (we use the name *Borel* MSO for the resulting logic):

A. The Borel MSO theory of $2^{\leq\omega}$ equipped with the prefix and lexicographic orders is decidable.
B. The Borel MSO theory of $\mathbb{R}$ with the usual order is decidable.

* Author supported by the Polish National Science Centre (NCN) grant "Polynomial finite state computation" (2022/46/A/ST6/00072).

† Partially supported by Deutsche Forschungsgemeinschaft (grant number 522843867) and European Research Council (grant number 101089343). Part of this work was done while Casares was at the University of Warsaw, Poland, supported by the Polish National Science Centre (NCN) grant "Polynomial finite state computation" (2022/46/A/ST6/00072).

‡ Author supported by Hausdorff Center for Mathematics at the University of Bonn (DFG EXC 2047).

§ Author supported by the Polish National Science Centre (NCN) grant "Frontiers of automatic analysis of concurrent systems" (2021/41/B/ST6/00535).

Conjecture B is the weaker one, since an isomorphic copy of the real numbers can be defined using Borel MSO in $2^{\leq\omega}$. In 2024, Manthe announced a proof of the weaker conjecture [7]. A proof of both conjectures was announced by Džamonja in 2025 [2], but we could not confirm its correctness.

This paper is devoted to the stronger Conjecture A. One can think of this conjecture as a common generalisation of Rabin's Tree Theorem and Conjecture B. This is because the complete binary tree is recovered in $2^{\leq\omega}$ as the definable substructure consisting only of nodes in 2^*. Hence, any proof of Conjecture A would also need to subsume a proof of Rabin's Tree Theorem. In particular, this seems to rule out the proof of Manthe, which is based on linear orders and the compositionality of their MSO theories, and does not seem to be amenable to extensions that cover the tree structure. This paper proposes a programme to prove Conjecture A. Our main contributions can be summarised as follows:

1. We introduce a certain kind of infinite games, and conjecture that these games can be won with finite memory.
2. We prove that if the conjecture on finite memory is true, then Conjecture A is also true, that is, one can decide the Borel MSO theory of $2^{\leq\omega}$, and furthermore the logic has a corresponding automaton model.
3. We present a preliminary study of the conjecture on finite memory, with some partial positive results.

Let us now describe the logics in question. We begin with the usual notion of monadic second-order logic (MSO). In this logic, there are two kinds of variables: lowercase variables x, y, z for elements and uppercase variables X, Y, Z for sets. Formulas are built using the following constructors:

$$\underbrace{\exists x\ \forall x}_{\substack{\text{quantification}\\ \text{over elements}}} \quad \underbrace{\exists X\ \forall X}_{\substack{\text{quantification}\\ \text{over sets}}} \quad \underbrace{\vee\ \wedge\ \neg}_{\substack{\text{Boolean}\\ \text{combinations}}} \quad \underbrace{R(x_1, \ldots, x_n)}_{\substack{\text{relations from}\\ \text{the vocabulary}}} \quad \underbrace{x \in X}_{\substack{\text{membership}\\ \text{relation}}}.$$

The semantics of the logic are defined in the usual way (see e.g. [13] for a more in-depth discussion). When we talk about the MSO theory of a structure, we mean the set of MSO sentences that are true in this structure. Rabin's Tree Theorem [9] states the decidability of the MSO theory of the infinite binary tree, which is viewed as a structure where the underlying set is 2^*, and there are two binary relations: the prefix order, and the lexicographic order.

The logic that is studied in this paper is *Borel* MSO. This logic has the same syntax as MSO, except that the semantics are changed so that set quantification is restricted to range only over Borel sets. Recall that the family of Borel sets in a topological space is the least family of subsets that contains the open sets, and which is closed under complementation as well as under taking countable unions and intersections. Therefore, in order to speak of Borel MSO, the underlying set in a structure must be equipped with a topology. The case which interests us is where the underlying set is $2^{\leq\omega}$, that is, strings of bits that have length at most ω. The finite strings are called *nodes* and the infinite ones are called *branches*. The topology is the one which arises taking as base of open sets the sets of the form

$$U_w = \{x \in 2^{\leq\omega} \mid w \text{ is a prefix of } x\}, \quad \text{where } w \in 2^*.$$

That is, the *open sets* of this topology are the (arbitrary) unions of U_w, for $w \in 2^*$. Note that the Borel restriction is relevant only for branches. This is because every set of nodes is Borel, and thus a subset of $2^{\leq\omega}$ is Borel if and only if its restriction to branches is a Borel set.

When we speak of $2^{\leq\omega}$ as a structure, we assume that it is equipped with the prefix and lexicographic orders, just as in Rabin's Tree Theorem. The lexicographic order gives us the tree structure on nodes, while the prefix relation allows us to check if some node belongs to a branch.

Example 1 (Determinacy). In this example, we illustrate the difference between MSO and Borel MSO. A Gale-Stewart game is a game where two players, call them Adam and Eve, pick bits from $2 = \{0, 1\}$ in alternation. After ω rounds, an infinite play in 2^ω arises. The winner of the game is given by a winning condition $W \subseteq 2^\omega$, which tells us which plays are winning for Eve. A strategy for player Eve can be described as a set of nodes X, which indicates the chosen successors used by the strategy. Similarly, a strategy for Adam can be described a set of nodes Y. Therefore, one can express determinacy of such games using a formula:

$$\forall W \subseteq 2^\omega.\ \big(\exists X \subseteq 2^*.\ \forall Y \subseteq 2^*.\ \varphi(W, X, Y) \quad \vee \quad \exists Y \subseteq 2^*.\ \forall X \subseteq 2^*.\ \neg\varphi(W, X, Y)\big),$$

where $\varphi(W, X, Y)$ says that the unique play consistent with strategies X and Y belongs to the set W. If we think of this formula as a formula of Borel MSO, then it is true, since only Borel winning conditions are considered, and Martin's Borel Determinacy Theorem says that such games are determined. If we think of this formula as a formula of (unrestricted) MSO, then it is false, since there exist games with (non-Borel) winning conditions that are not determined, assuming the Axiom of Choice.

This paper is devoted to Conjecture A mentioned above, which says that the Borel MSO theory of $2^{\leq\omega}$ is decidable, meaning that there is an algorithm that inputs a sentence of the logic, and answers whether the sentence is true in the structure. Simply put, our programme is to copy the proof of Rabin's Tree Theorem, and extend it to take into account Borel sets of branches.

We begin by briefly recalling how Rabin's Tree Theorem is proved, using a modern presentation that emphasises games, as proposed by Gurevich and Harrington [5]. In the proof, one shows that every MSO formula can be transformed into an equivalent automaton, and the latter can be tested for emptiness. Proving this requires to show that automata have the same closure properties as the logic, namely Boolean combinations and projection. Union and intersection are easily achieved using a product construction, and projection is also easy because the automaton model in question is nondeterministic. The difficult closure property is complementation. Unlike for ω-words, determinisation is not an option. To prove complementation, one describes rejection by the automaton in terms of a game of infinite duration, and then proves that this game can be won using simple strategies of finite memory. The finite memory can then be embedded into the state space of a nondeterministic automaton that recognises the complementary language.

The key message of this paper is that most of the proof discussed in the previous paragraph can be adapted to the more general setting of $2^{\leq\omega}$. In this more general setting, an input tree has type $2^{\leq\omega} \to \Sigma$, that is, it assigns labels to both nodes and branches, and a run of the automaton has type $2^{\leq\omega} \to Q$. In both cases, the maps are required to be Borel, since on the logical side we are targeting Borel MSO. The acceptance condition for the automaton is a variant of the Muller condition that takes into account the states assigned to the branches. Using this automaton model, we are able to implement all parts of the proof of Rabin's Tree Theorem, with one exception, namely finite-memory determinacy of the underlying games. This determinacy result is the content of our conjecture, and much of the paper is devoted to its study.

Let us now describe in more detail our conjecture on finite-memory strategies. Like most games studied in the context of automata, such as parity games or Muller games, our games are infinite duration games played by two players, called Adam and Eve, using directed graphs (called arenas), where each position is owned by one of the two players. The definition below highlights only the features that distinguish games arising in the context of Borel MSO from the usual notion of games. We call the games Muller-Borel, because the acceptance condition is a Boolean combination of Büchi and Borel conditions, and Boolean combinations of Büchi conditions are called Muller conditions.

Definition 2. *A* Muller-Borel game *is a game satisfying the following two conditions:*

(1) ***Tree-like arena.*** *The set of positions is $2^* \times Q$ for a finite set Q. Edges of the game can only move to child nodes, that is, if*

$$(v, q) \to (w, p)$$

is an edge in the game, then w is a child of v. Not all edges that move to children need to be available.

(2) ***Muller-Borel condition.*** *The winning condition is a finite Boolean combination of two kinds of conditions:*

(Büchi) *$q \in Q$ is seen infinitely often;*

(Borel) *the branch belongs to a Borel set $B \subseteq 2^\omega$.*

Using Martin's Borel Determinacy Theorem [8, p. 371], one can show that the games are determined, that is, one of the players must have a winning strategy. However, our main interest is in the memory used by the winning strategy, which is beyond the scope of Martin's theorem. Our conjecture is that the games can be won with finite memory, as stated below.

Conjecture 3. *For every Muller-Borel game, one of the players has a winning strategy that uses finite memory. Furthermore, an upper bound on the memory size can be computed based on the size of Q.*

Using the same framework as for Rabin's Tree Theorem, we prove that the above conjecture implies decidability of Borel MSO on $2^{\leq\omega}$, that is, Shelah's Conjecture A. The rest of this abstract is devoted to a preliminary discussion of the conjecture.

As indicated, the main examples of Muller-Borel games appear in the analysis of tree automata. One can recognize these games in the tree-like requirement (Definition 2, (1)), with Q being the set of states. For many kinds of games, such as Muller games or parity games, the tree-like shape of the arena is irrelevant in the analysis of the game. Hence, it is common to define those games just over arbitrary graphs [12, 3]. In our case, however, the tree-like shape is essential, since the Borel conditions in the game are defined in terms of the tree structure. Let us now explain how certain special cases of a Muller-Borel game correspond to important results on determinacy of games.

- Suppose that we only allow Büchi conditions in the game. In this case, we have a game with a Muller condition. Muller games are known to have finite-memory strategies, which was first proved by Büchi and Landweber for finite arenas [1, Theorem 1'], and then by Gurevich and Harrington for general arenas [5, Theorem 1]. Therefore, if we only allow Büchi conditions, then the conjecture is true.
- Suppose that there is only one state in the game. In this case, each game position can be reached in only one way (since the positions form a tree), and therefore all strategies are memoryless. Thus, from Martin's Borel Determinacy Theorem we can conclude that the conjecture holds in this special case, even without any memory. Note that the full power of Martin's theorem is needed for this special case.

In the paper, we prove the conjecture for other special cases of the game, including the special case where only the Borel condition is allowed (which extends the situation from the second item above, in which the Büchi condition trivializes).

Our conjecture talks about finite-memory strategies. A natural question would be about memoryless strategies. For example, in automata theory, finite-memory determinacy of Muller games is commonly proved by first reducing to a parity condition, and then proving that parity games can be won using memoryless strategies. The only need for memory arises from the reduction from Muller to parity, which is typically accomplished using a variant of McNaughton's "latest appearance record" construction. For our conjecture, we are unable to present such a reduction, that is, we are not aware of a variant of our conjecture which would talk about memoryless strategies. Using priorities does not seem to help: memory is needed even if we convert the game into a parity-style format, in which the Büchi part of the acceptance condition is replaced by a priority function. In fact, even in the presence of only three priorities, the amount of needed memory grows unboundedly as a function of the number of states. On the positive side, we show that for two priorities, only two states of memory are needed, regardless of the number of states in Q. We also prove a few other lower and upper bounds.

The conjecture seems rather strong. In particular, the conjecture implies that if we restrict the semantics of Borel MSO to use only Boolean combinations of F_σ sets (i.e., of countable unions of closed sets), then the same formulae are true. In other words, Borel MSO can essentially only talk about Boolean combinations of F_σ sets, and it cannot access more complicated Borel sets. Since one could expect Borel MSO to be more expressive, this could be seen as evidence against the conjecture. However, the recent proof of Manthe is consistent with our conjecture, since it implies that in the weaker setting of 2^ω one can also restrict quantification to Boolean combinations of F_σ sets, and the same formulae are true.

[1] J. Richard Büchi and Lawrence H. Landweber. Solving sequential conditions by finite-state strategies. *Transactions of the American Mathematical Society*, 138:295–311, 1969.

[2] Mirna Džamonja. MSO logic of the real order with the set quantifiers ranging over the Borel sets. 2025.

[3] Nathanaël Fijalkow, C. Aiswarya, Guy Avni, Nathalie Bertrand, Patricia Bouyer, Romain Brenguier, Arnaud Carayol, Antonio Casares, John Fearnley, Paul Gastin, Hugo Gimbert, Thomas A. Henzinger, Florian Horn, Rasmus Ibsen-Jensen, Nicolas Markey, Benjamin Monmege, Petr Novotný, Pierre Ohlmann, Mickael Randour, Ocan Sankur, Sylvain Schmitz, Olivier Serre, Mateusz Skomra, Nathalie Sznajder, and Pierre Vandenhove. Games on graphs: From logic and automata to algorithms, 2025. To appear in Cambridge University Press.

[4] Yuri Gurevich. Monadic second-order theories. In *Model-theoretic logics*, Perspectives in Mathematical Logic, pages 479–506. Springer, New York, 1985.

[5] Yuri Gurevich and Leo Harrington. Trees, automata and games. In *Proc. ACM Symp. on Theory of Computing*, pages 60–65. Assoc. Comput. Mach., 1982.

[6] Yuri Gurevich and Saharon Shelah. Monadic theory of order and topology in ZFC. *Ann. Math. Logic*, 23(2-3):179–198, 1982.

[7] Sven Manthe. The Borel monadic theory of order is decidable. 2024.

[8] Donald A. Martin. Borel determinacy. *Ann. of Math. (2)*, 102(2):363–371, 1975.

[9] Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.

[10] Saharon Shelah. The monadic theory of order. *Ann. of Math. (2)*, 102(3):379–419, 1975.

[11] Saharon Shelah. On what I do not understand (and have something to say), model theory. *Math. Japon.*, 51(2):329–377, 2000.

[12] Wolfgang Thomas. On the synthesis of strategies in infinite games. In *Proc. STACS 95*, volume 900 of *LNCS*, pages 1–13. Springer, 1995.

[13] Wolfgang Thomas. Languages, automata, and logic. In *Handbook of formal languages, Vol. 3*, pages 389–455. Springer, Berlin, 1997.