IST AUSTRIA
Institute of Science and Technology

What is Decidable about Partially Observable Markov Decision Processes with -Regular Objectives

Krishnendu Chatterjee and Martin Chmelík and Tracol Mathieu

What is Decidable about Partially Observable Markov Decision Processes with ω-Regular Objectives (Full Version)

Krishnendu Chatterjee   Martin Chmelik   Mathieu Tracol

(IST Austria)

Abstract

We consider partially observable Markov decision processes (POMDPs) with ω-regular conditions specified as parity objectives. The class of ω-regular languages extends regular languages to infinite strings and provides a robust specification language to express all properties used in verification, and parity objectives are canonical forms to express ω-regular conditions. The qualitative analysis problem given a POMDP and a parity objective asks whether there is a strategy to ensure that the objective is satisfied with probability 1 (resp. positive probability). While the qualitative analysis problems are known to be undecidable even for very special cases of parity objectives, we establish decidability (with optimal complexity) of the qualitative analysis problems for POMDPs with all parity objectives under finite-memory strategies. We establish asymptotically optimal (exponential) memory bounds and EXPTIME-completeness of the qualitative analysis problems under finite-memory strategies for POMDPs with parity objectives.

1 Introduction

Partially observable Markov decision processes (POMDPs). *Markov decision processes (MDPs)* are standard models for probabilistic systems that exhibit both probabilistic and nondeterministic behavior [19]. MDPs have been used to model and solve control problems for stochastic systems [16]: nondeterminism represents the freedom of the controller to choose a control action, while the probabilistic component of the behavior describes the system response to control actions. In *perfect-observation (or perfect-information) MDPs (PIMDPs)* the controller can observe the current state of the system to choose the next control actions, whereas in *partially observable MDPs (POMDPs)* the state space is partitioned according to observations that the controller can observe i.e., given the current state, the controller can only view the observation of the state (the partition the state belongs to), but not the precise state [27]. POMDPs provide the appropriate model to study a wide variety of applications such as in computational biology [15], speech processing [26], image processing [13], software verification [6], robot planning [22], reinforcement learning [20], to name a few. In verification of probabilistic systems, MDPs have been adopted as models for concurrent probabilistic systems [12], probabilistic systems operating in open environments [33], under-specified probabilistic systems [4], and applied in diverse domains [3, 24]. POMDPs also subsume many other powerful computational models such as probabilistic automata [30, 28] (since probabilistic automata (aka blind POMDPS) are a special case of POMDPs where there is only a single observation).

The class of ω-regular objectives. An objective specifies the desired set of behaviors (or paths) for the controller. In verification and control of stochastic systems an objective is typically an ω-regular set of paths. The class of ω-regular languages extends classical regular languages to infinite strings, and provides a robust specification language to express all commonly used specifications, such as safety, reachability, liveness, fairness, etc [35]. In a parity objective, every state of the MDP is mapped to a non-negative integer priority (or color) and the goal is to ensure that the minimum priority visited infinitely often is even. Parity objectives are a canonical way to define such ω-regular specifications (e.g., all specifications in verification expressed as a linear-time temporal logic (LTL) formula can be translated to a parity objective). Thus POMDPs with parity objectives provide the theoretical framework to study problems such as the verification and control of stochastic systems.

Qualitative and quantitative analysis. The analysis of POMDPs with parity objectives can be classified into qualitative and quantitative analysis. Given a POMDP with a parity objective and a start state, the *qualitative analysis* asks whether the objective can be ensured with probability 1 (*almost-sure winning*) or positive probability (*positive winning*); whereas the *quantitative analysis* asks whether the objective can be satisfied with probability at least λ for a given threshold $\lambda \in (0, 1)$.

Importance of qualitative analysis. The qualitative analysis of MDPs is an important problem in verification that is of interest independent of the quantitative analysis problem. There are many applications where we need to know whether the correct behavior arises with probability 1. For instance, when analyzing a randomized embedded scheduler, we are interested in whether every thread progresses with probability 1 [14]. Even in settings where it suffices to satisfy certain specifications with probability $\lambda < 1$, the correct choice of λ is a challenging problem, due to the simplifications introduced during modeling. For example, in the analysis of randomized distributed algorithms it is quite common to require correctness with probability 1 (see, e.g., [29, 23, 34]). Furthermore, in contrast to quantitative analysis, qualitative analysis is robust to numerical perturbations and modeling errors in the transition probabilities. Thus qualitative analysis of POMDPs with parity objectives is one of the most fundamental theoretical problems in verification and analysis of probabilistic systems.

Previous results. On one hand POMDPs with parity objectives provide a rich framework to model a wide variety of practical problems, on the other hand, most theoretical results established for POMDPs are *negative* (undecidability) results. There are several deep undecidability results established for the special case of probabilistic automata (that immediately imply undecidability for the more general case of POMDPs). The basic undecidability results are for probabilistic automata over finite words (that can be considered as a special case of parity objectives). The quantitative analysis problem is undecidable for probabilistic automata over finite words [30, 28, 11]; and it was shown in [25] that even the following approximation version is undecidable: for any fixed $0 < \epsilon < \frac{1}{2}$, given a probabilistic automaton and the guarantee that either (a) all words are accepted with probability at least $1 - \epsilon$; or (ii) all words are accepted with probability at most ϵ; decide whether it is case (i) or case (ii). The almost-sure (resp. positive) problem for probabilistic automata over finite words reduces to the non-emptiness question of universal (resp. non-deterministic) automata over finite words and is PSPACE-complete (resp. solvable in polynomial time). However, another related decision question whether for every $\epsilon > 0$ there is a word that is accepted with probability at least $1 - \epsilon$ (called the value 1 problem) is undecidable for probabilistic automata over finite words [17]. Also observe that all undecidability results for probabilistic automata over finite words carry over to POMDPs where the controller is restricted to finite-memory strategies. For qualitative analysis of POMDPs with parity objectives, deep undecidability results were established even for very special cases of parity objectives (even in the special case of probabilistic automata). It was shown in [1, 2] that the almost-sure (resp. positive) problem is undecidable for probabilistic automata with coBüchi (resp. Büchi) objectives which are special cases of

parity objectives that use only two priorities. In summary the most important theoretical results are negative in the sense that they establish undecidability results.

Our contributions. For POMDPs with parity objectives, all questions related to quantitative analysis are undecidable, and the qualitative analysis problems are also undecidable in general. However, the undecidability proofs for the qualitative analysis of POMDPs with parity objectives crucially require the use of *infinite-memory* strategies for the controller. In all practical applications, the controller must be a *finite-state* controller to be implementable. Thus for all practical purposes the relevant question is the existence of finite-memory controllers. The quantitative analysis problem remains undecidable even under finite-memory controllers as the undecidability results are established for probabilistic automata over finite words. In this work we study the most prominent remaining theoretical open question (that is also of practical relevance) for POMDPs with parity objectives that whether the qualitative analysis of POMDPs with parity objectives is decidable or undecidable for finite-memory strategies (i.e., finite-memory controllers). Our main result is the *positive* result that qualitative analysis of POMDPs with parity objectives is *decidable* under finite-memory strategies. Moreover, for qualitative analysis of POMDPs with parity objectives under finite-memory strategies we establish asymptotically optimal complexity bounds both for strategy complexity as well as computational complexity. The details of our contributions are as follows:

1. *(Strategy complexity).* Our first result shows that *belief-based* strategies are not sufficient (where a belief-based strategy is based on the subset construction that remembers the possible set of current states): we show that there exist POMDPs with coBüchi objectives where finite-memory almost-sure winning strategy exists but there exists no randomized belief-based almost-sure winning strategy. All previous results about decidability for almost-sure winning in sub-classes of POMDPs crucially relied on the sufficiency of randomized belief-based strategies that allowed standard techniques like subset construction to establish decidability. However, our counter-example shows that previous techniques based on simple subset construction (to construct an exponential size PIMDP) are not adequate to solve the problem. Before the result for parity objectives, we consider a slightly more general form of objectives, called Muller objectives. For a Muller objective a set $\mathcal{F}$ of subsets of colors is given and the set of colors visited infinitely often must belong to $\mathcal{F}$. We show our main result that given a POMDP with $|S|$ states and a Muller objective with d colors (priorities), if there is a finite-memory almost-sure (resp. positive) winning strategy, then there is an almost-sure (resp. positive) winning strategy that uses at most $\mathsf{Mem}^* = 2^{2 \cdot |S|} \cdot (2^{2^d})^{|S|}$ memory. Developing on our result for Muller objectives, for POMDPs with parity objectives we show that if there is a finite-memory almost-sure (resp. positive) winning strategy, then there is an almost-sure (resp. positive) winning strategy that uses at most $2^{3 \cdot d \cdot |S|}$ memory. Our exponential memory upper bound for parity objectives is asymptotically optimal as it has been already established in [8] that almost-sure winning strategies require at least exponential memory even for the very special case of reachability objectives in POMDPs.

2. *(Computational complexity).* We present an exponential time algorithm for the qualitative analysis of POMDPs with parity objectives under finite-memory strategies, and thus obtain an EXPTIME upper bound. The EXPTIME-hardness follows from [8] for the special case of reachability and safety objectives, and thus we obtain the optimal EXPTIME-complete computational complexity result.

In Table 1 and Table 2 we summarize the results for strategy complexity and computational complexity, respectively.

Technical contributions. The key technical contribution for the decidability result is as follows. Since belief-based strategies are not sufficient, standard subset construction techniques do not work. For an arbitrary

finite-memory strategy we construct a projected strategy that collapses memory states based on a projection graph construction given the strategy. The projected strategy at a collapsed memory state plays uniformly over actions that were played at all the corresponding memory states of the original strategy. The projected strategy thus plays more actions with positive probability. The key challenge is to show the bound on the size of the projection graph, and to show that the projected strategy, even though plays more actions, does not destroy the structure of the recurrent classes of the original strategy. For parity objectives, we show a reduction from general parity objectives to parity objectives with two priorities on a polynomially larger POMDP and from our general result for Muller objectives obtain the asymptotically optimal memory complexity bounds for parity objectives. For the computational complexity result, we show how to construct an exponential size special class of POMDPs (which we call belief-observation POMDPs where the belief is always the current observation) and present polynomial time algorithms for the qualitative analysis of the special belief-observation POMDPs of our construction.

Objectives	Almost-sure		Positive		Quantitative	
	Inf. Mem.	Fin. Mem.	Inf. Mem.	Fin. Mem.	Inf. Mem.	Fin. Mem.
Büchi	Exp. (belief-based)	Exp. (belief-based)	Inf. mem. req.	UB: **Exp. $2^{6\cdot|S|}$** LB: Exp. **(belief not sufficient)**	Inf. mem. req.	No bnd.
coBüchi	Inf. mem. req.	UB: **Exp. $2^{6\cdot|S|}$** LB: Exp. **(belief not sufficient)**	UB: Exp. LB: Exp. **(belief not sufficient)**	UB: Exp. LB: Exp. **(belief not sufficient)**	Inf. mem. req.	No bnd.
Parity	Inf. mem. req.	UB: **Exp. $2^{3\cdot d\cdot|S|}$** LB: Exp. **(belief not sufficient)**	Inf. mem. req.	UB: **Exp. $2^{3\cdot d\cdot|S|}$** LB: Exp. **(belief not sufficient)**	Inf. mem. req.	No bnd.

Table 1: Strategy complexity for POMDPs with parity objectives, where $|S|$ is the size of state space, and d the number of priorities, (UB denotes upper bound and LB denotes lower bound). The results in boldface are new results included in the present paper.

Objectives	Almost-sure		Positive		Quantitative	
	Inf. Mem.	Finite Mem.	Inf. Mem.	Finite Mem.	Inf. Mem.	Finite Mem.
Büchi	EXP-complete	EXP-complete	Undec.	**EXP-complete**	Undec.	Undec.
coBüchi	Undec.	**EXP-complete**	EXP-complete	EXP-complete	Undec.	Undec.
Parity	Undec.	**EXP-complete**	Undec.	**EXP-complete**	Undec.	Undec.

Table 2: Computational complexity for POMDPs with parity objectives. The results in boldface are new results included in the present paper.

2 Definitions

In this section we present the basic definitions of POMDPs, strategies (policies), ω-regular objectives, and the winning modes.

Notations. Given a finite set X, we denote by $\mathcal{P}(X)$ the set of subsets of X, i.e., $\mathcal{P}(X)$ is the power set of X. A probability distribution f on X is a function $f : X \to [0,1]$ such that $\sum_{x \in X} f(x) = 1$, and we denote by $\mathcal{D}(X)$ the set of all probability distributions on X. For $f \in \mathcal{D}(X)$ we denote by $\mathrm{Supp}(f) = \{x \in X \mid f(x) > 0\}$ the support of f.

Definition 1 (POMDP). *A* Partially Observable Markov Decision Process (POMDP) *is a tuple* $G = (S, A, \delta, \mathcal{O}, \gamma, s_0)$ *where:*

- S *is a finite set of states;*

- A *is a finite alphabet of* actions*;*

- $\delta : S \times A \to \mathcal{D}(S)$ *is a* probabilistic transition function *that given a state s and an action $a \in A$ gives the probability distribution over the successor states, i.e., $\delta(s,a)(s')$ denotes the transition probability from state s to state s' given action a;*

- $\mathcal{O}$ *is a finite set of* observations*;*

- $\gamma : S \to \mathcal{O}$ *is an* observation function *that maps every state to an observation; and*

- s_0 *is the initial state.*

Given $s, s' \in S$ and $a \in A$, we also write $\delta(s'|s,a)$ for $\delta(s,a)(s')$. For an observation o, we denote by $\gamma^{-1}(o) = \{s \in S \mid \gamma(s) = o\}$ the set of states with observation o. For a set $U \subseteq S$ of states and $O \subseteq \mathcal{O}$ of observations we denote $\gamma(U) = \{o \in \mathcal{O} \mid \exists s \in U.\ \gamma(s) = o\}$ and $\gamma^{-1}(O) = \bigcup_{o \in O} \gamma^{-1}(o)$.

Remark 1. *For technical convenience we have assumed that there is an unique initial state and we will also assume that the initial state s_0 has a unique observation, i.e., $|\gamma^{-1}(\gamma(s_0))| = 1$. In general there is an initial distribution α over initial states that all have the same observation, i.e., $\mathrm{Supp}(\alpha) \subseteq \gamma^{-1}(o)$, for some $o \in \mathcal{O}$. However, this can be modeled easily by adding a new initial state s_{new} with a unique observation such that in the first step gives the desired initial probability distribution α, i.e., $\delta(s_{new}, a) = \alpha$ for all actions $a \in A$. Hence for simplicity we assume there is a unique initial state s_0 with a unique observation.*

Plays, cones and belief-updates. A *play* (or a path) in a POMDP is an infinite sequence $(s_0, a_0, s_1, a_1, s_2, a_2, \ldots)$ of states and actions such that for all $i \geq 0$ we have $\delta(s_i, a_i)(s_{i+1}) > 0$. We write Ω for the set of all plays. For a finite prefix $w \in (S \cdot A)^* \cdot S$ of a play, we denote by $\mathsf{Cone}(w)$ the set of plays with w as the prefix (i.e., the cone or cylinder of the prefix w), and denote by $\mathsf{Last}(w)$ the last state of w. For a finite prefix $w = (s_0, a_0, s_1, a_1, \ldots, s_n)$ we denote by $\gamma(w) = (\gamma(s_0), a_0, \gamma(s_1), a_1, \ldots, \gamma(s_n))$ the observation and action sequence associated with w. For a finite sequence $\rho = (o_0, a_0, o_1, a_1, \ldots, o_n)$ of observations and actions, the *belief* $\mathcal{B}(\rho)$ after the prefix ρ is the set of states in which a finite prefix of a play can be after the sequence ρ of observations and actions, i.e., $\mathcal{B}(\rho) = \{s_n = \mathsf{Last}(w) \mid w = (s_0, a_0, s_1, a_1, \ldots, s_n), w$ is a prefix of a play, and for all $0 \leq i \leq n.\ \gamma(s_i) = o_i\}$. The belief-updates associated with finite-prefixes are as follows: for prefixes w and $w' = w \cdot a \cdot s$ the belief update is defined inductively as $\mathcal{B}(\gamma(w')) = \left(\bigcup_{s_1 \in \mathcal{B}(\gamma(w))} \mathrm{Supp}(\delta(s_1, a)) \right) \cap \gamma^{-1}(s)$, i.e., the set $\left(\bigcup_{s_1 \in \mathcal{B}(\gamma(w))} \mathrm{Supp}(\delta(s_1, a)) \right)$

5

denotes the possible successors given the belief $\mathcal{B}(\gamma(w))$ and action a, and then the intersection with the set of states with the current observation $\gamma(s)$ gives the new belief set.

Strategies. A *strategy (or a policy)* is a recipe to extend prefixes of plays and is a function $\sigma : (S \cdot A)^* \cdot S \to \mathcal{D}(A)$ that given a finite history (i.e., a finite prefix of a play) selects a probability distribution over the actions. Since we consider POMDPs, strategies are *observation-based*, i.e., for all histories $w = (s_0, a_0, s_1, a_1, \ldots, a_{n-1}, s_n)$ and $w' = (s'_0, a_0, s'_1, a_1, \ldots, a_{n-1}, s'_n)$ such that for all $0 \leq i \leq n$ we have $\gamma(s_i) = \gamma(s'_i)$ (i.e., $\gamma(w) = \gamma(w')$), we must have $\sigma(w) = \sigma(w')$. In other words, if the observation sequence is the same, then the strategy cannot distinguish between the prefixes and must play the same. We now present an equivalent definition of observation-based strategies such that the memory of the strategy is explicitly specified, and will be required to present finite-memory strategies.

Definition 2 (Strategies with memory and finite-memory strategies)**.** *A strategy* with memory *is a tuple* $\sigma = (\sigma_u, \sigma_n, M, m_0)$ *where:*

- (Memory set). M *is a denumerable set (finite or infinite) of memory elements (or memory states).*

- (Action selection function). *The function* $\sigma_n : M \to \mathcal{D}(A)$ *is the* action selection function *that given the current memory state gives the probability distribution over actions.*

- (Memory update function). *The function* $\sigma_u : M \times \mathcal{O} \times A \to \mathcal{D}(M)$ *is the* memory update function *that given the current memory state, the current observation and action, updates the memory state probabilistically.*

- (Initial memory). *The memory state* $m_0 \in M$ *is the initial memory state.*

A strategy is a finite-memory *strategy if the set* M *of memory elements is finite. A strategy is* pure (or deterministic) *if the memory update function and the action selection function are deterministic, i.e.,* $\sigma_u : M \times \mathcal{O} \times A \to M$ *and* $\sigma_n : M \to A$. *It was shown in [7] that in POMDPs pure strategies are as powerful as randomized strategies, hence in sequel we omit discussions about pure strategies. A strategy is* memoryless (or stationary) *if it is independent of the history but depends only on the current observation, and can be represented as a function* $\sigma : \mathcal{O} \to \mathcal{D}(A)$.

Probability measure. Given a strategy σ, the unique probability measure obtained given σ is denoted as $\mathbb{P}^\sigma(\cdot)$. We first define the measure $\mu^\sigma(\cdot)$ on cones. For $w = s_0$ we have $\mu^\sigma(\mathsf{Cone}(w)) = 1$, and for $w = s$ where $s \neq s_0$ we have $\mu^\sigma(\mathsf{Cone}(w)) = 0$; and for $w' = w \cdot a \cdot s$ we have $\mu^\sigma(\mathsf{Cone}(w')) = \mu^\sigma(\mathsf{Cone}(w)) \cdot \sigma(w)(a) \cdot \delta(\mathsf{Last}(w), a)(s)$. By Caratheódary's extension theorem, the function $\mu^\sigma(\cdot)$ can be uniquely extended to a probability measure $\mathbb{P}^\sigma(\cdot)$ over Borel sets of infinite plays [5].

Objectives. An *objective* in a POMDP G is a Borel set $\varphi \subseteq \Omega$ of plays in the Cantor topology on Ω [21]. All objectives we consider in this paper lie in the first $2\,1/2$-levels of the Borel hierarchy. We specifically consider the parity objective, which is a canonical form to express all ω-regular objectives [35]. Thus parity objectives provide a robust specification language to express all commonly used properties in verification and system analysis. For a play $\rho = (s_0, a_0, s_1, a_1, s_2 \ldots)$, we denote by $\mathsf{Inf}(\rho) = \{s \in S \mid \forall i \geq 0 \cdot \exists j \geq i : s_j = s\}$ the set of states that occur infinitely often in ρ. We consider the following objectives.

- *Reachability and safety objectives.* Given a set $\mathcal{T} \subseteq S$ of target states, the *reachability* objective $\mathsf{Reach}(\mathcal{T}) = \{(s_0, a_0, s_1, a_1, s_2 \ldots) \in \Omega \mid \exists k \geq 0 : s_k \in \mathcal{T}\}$ requires that a target state in $\mathcal{T}$ is visited at least once. Dually, the *safety* objective $\mathsf{Safe}(\mathcal{T}) = \{(s_0, a_0, s_1, a_1, s_2 \ldots) \in \Omega \mid \forall k \geq 0 : s_k \in \mathcal{T}\}$ requires that only states in $\mathcal{T}$ are visited.

- *Büchi and coBüchi objectives.* Given a set $\mathcal{T} \subseteq S$ of target states, the *Büchi* objective $\mathsf{Buchi}(\mathcal{T}) = \{\rho \in \Omega \mid \mathrm{Inf}(\rho) \cap \mathcal{T} \neq \emptyset\}$ requires that a state in $\mathcal{T}$ is visited infinitely often. Dually, the *coBüchi* objective $\mathsf{coBuchi}(\mathcal{T}) = \{\rho \in \Omega \mid \mathrm{Inf}(\rho) \subseteq \mathcal{T}\}$ requires that only states in $\mathcal{T}$ are visited infinitely often.

- *Parity objectives.* For $d \in \mathbb{N}$, let $p : S \to \{0, 1, \ldots, d\}$ be a *priority function* that maps each state to a non-negative integer priority. The *parity* objective $\mathsf{Parity}(p) = \{\rho \in \Omega \mid \min\{p(s) \mid s \in \mathrm{Inf}(\rho)\}$ is even$\}$ requires that the smallest priority that appears infinitely often is even.

- *Muller objectives.* Let D be a set of colors, and $\mathsf{col} : S \to D$ be a color mapping function that maps every state to a color. A Muller objective $\mathcal{F}$ consists of a set of subsets of colors and requires that the set of colors visited infinitely often belongs to $\mathcal{F}$, i.e., $\mathcal{F} \in \mathcal{P}(\mathcal{P}(D))$ and $\mathsf{Muller}(\mathcal{F}) = \{\rho \in \Omega \mid \{\mathsf{col}(s) \mid s \in \mathrm{Inf}(\rho)\} \in \mathcal{F}\}$

Note that a reachability objective $\mathsf{Reach}(\mathcal{T})$ can be viewed as a special case of Büchi as well as coBüchi objectives, (assuming w.l.o.g. that all target states $s \in \mathcal{T}$ are *absorbing*, i.e., $\delta(s, a)(s) = 1$ for all $a \in A$) and analogously safety objectives are also special cases of Büchi and coBüchi objectives. The objectives $\mathsf{Buchi}(\mathcal{T})$ and $\mathsf{coBuchi}(\mathcal{T})$ are special cases of parity objectives defined by respective priority functions p_1, p_2 such that $p_1(s) = 0$ and $p_2(s) = 2$ if $s \in \mathcal{T}$, and $p_1(s) = p_2(s) = 1$ otherwise. Given a set $U \subseteq S$ we will denote by $p(U)$ the set of priorities of the set U given by the priority function p, i.e., $p(U) = \{p(s) \mid s \in U\}$, and similarly $\mathsf{col}(U) = \{\mathsf{col}(s) \mid s \in U\}$. Also observe that parity objectives are a special case of Muller objectives, however, given a POMDP with a Muller objective with color set D, an equivalent POMDP with $|S| \cdot |D|!$ states and a parity objective with $|D|^2$ priorities can be constructed using the well-known latest appearance record (LAR) construction of [18] for conversion of Muller objectives to parity objectives. An objective φ is visible if for all plays ρ and ρ' that have the same observation sequence we have $\rho \in \varphi$ iff $\rho' \in \varphi$.

Winning modes. Given a POMDP, an objective φ, and a class $\mathcal{C}$ of strategies, we say that:

- a strategy $\sigma \in \mathcal{C}$ is *almost-sure winning* if $\mathbb{P}^\sigma(\varphi) = 1$;

- a strategy $\sigma \in \mathcal{C}$ is *positive winning* if $\mathbb{P}^\sigma(\varphi) > 0$;

- the POMDP is *limit-sure winning* if for all $\varepsilon > 0$ there exists a strategy $\sigma \in \mathcal{C}$ for player 1 such that $\mathbb{P}^\sigma(\varphi) \geq 1 - \epsilon$; and

- a strategy $\sigma \in \mathcal{C}$ is *quantitative winning*, for a threshold $\lambda \in (0, 1)$, if $\mathbb{P}^\sigma(\varphi) \geq \lambda$.

Theorem 1 (Decidability and complexity under general strategies [30, 28, 11, 17, 10, 1, 2, 31, 32, 8])**.** *The following assertions hold for POMDPs with the class $\mathcal{C}$ of all infinite-memory (randomized or pure) strategies:*

1. *The quantitative winning problem is undecidable for safety, reachability, Büchi, coBüchi, parity, and Muller objectives.*

2. *The limit-sure winning problem is EXPTIME-complete for safety objectives; and undecidable for reachability, Büchi, coBüchi, parity, and Muller objectives.*

3. *The almost-sure winning problem is EXPTIME-complete for safety, reachability, and Büchi objectives; and undecidable for coBüchi, parity, and Muller objectives.*

4. *The positive winning problem is PTIME-complete for reachability objectives, EXPTIME-complete for safety and coBüchi objectives; and undecidable for Büchi, parity, and Muller objectives.*

Explanation of the previous results and implications under finite-memory policies. All the undecidability results follow from the special case of probabilistic automata: the undecidability of the quantitative problem for probabilistic automata follows from [30, 28, 11]; the undecidability of the limit-sure winning for finite words and reachability objectives was established in [17, 10] (the undecidability of limit-sure reachability also implies undecidability for Büchi, coBüchi and parity objectives); the undecidability for positive winning for Büchi and almost-sure winning for coBüchi objectives was established in [1, 2]. For the decidable results, the optimal complexity results for safety objectives can be obtained from the results of [31, 32] and all the other results follow from [8, 2]. If the classes of strategies are restricted to finite-memory strategies, then the undecidability results for quantitative winning and limit-sure winning still hold, as they are established for reachability objectives and for reachability objectives finite-memory suffices. The most prominent and important open question is whether the almost-sure and positive winning problems are decidable for parity and Muller objectives in POMDPs under finite-memory strategies. All the lower bounds (i.e., hardness and undecidability) results have been established for the cases when the objectives are restricted to be visible.

3 Strategy Complexity for Muller Objectives under Finite-memory Strategies

In this section we will first show that randomized belief-based strategies are not sufficient for finite-memory almost-sure winning strategies in POMDPs with coBüchi objectives; and then present the upper bound on memory size required for finite-memory almost-sure and positive winning strategies in POMDPs with Muller objectives. Our proofs will use many basic results on Markov chains and we start with them in the following subsection.

3.1 Basic properties of Markov chains

Since our proof relies heavily on Markov chains we start with some basic definitions and properties related to Markov chains that are essential for our proofs.

Markov chains, recurrent classes, and reachability. A Markov chain $\overline{G} = (\overline{S}, \overline{\delta})$ consists of a finite set $\overline{S}$ of states and a probabilistic transition function $\overline{\delta} : \overline{S} \to \mathcal{D}(\overline{S})$. Given the Markov chain, we consider the directed graph $(\overline{S}, \overline{E})$ where $\overline{E} = \{(\overline{s}, \overline{s}') \mid \delta(\overline{s}' \mid \overline{s}) > 0\}$. A *recurrent class* $\overline{C} \subseteq \overline{S}$ of the Markov chain is a bottom strongly connected component (scc) in the graph $(\overline{S}, \overline{E})$ (a bottom scc is an scc with no edges out of the scc). We denote by $\mathsf{Rec}(\overline{G})$ the set of recurrent classes of the Markov chain, i.e., $\mathsf{Rec}(\overline{G}) = \{\overline{C} \mid \overline{C} \text{ is a recurrent class}\}$. Given a state $\overline{s}$ and a set $\overline{U}$ of states, we say that $\overline{U}$ is reachable from $\overline{s}$ if there is a path from $\overline{s}$ to some state in $\overline{U}$ in the graph $(\overline{S}, \overline{E})$. Given a state $\overline{s}$ of the Markov chain we denote by $\mathsf{Rec}(\overline{G})(\overline{s}) \subseteq \mathsf{Rec}(\overline{G})$ the subset of the recurrent classes reachable from $\overline{s}$ in $\overline{G}$. A state is recurrent if it belongs to a recurrent class. The following standard properties of reachability and the recurrent classes will be used in our proof:

1. *Property 1.* (a) For a set $\overline{T} \subseteq \overline{S}$, if for all states $\overline{s} \in \overline{S}$ there is a path to $\overline{T}$ (i.e., for all states there is a positive probability to reach $\overline{T}$), then from all states the set $\overline{T}$ is reached with probability 1. (b) For all states $\overline{s}$, if the Markov chain starts at $\overline{s}$, then the set $\mathcal{C} = \bigcup_{\overline{C} \in \mathsf{Rec}(\overline{G})(\overline{s})} \overline{C}$ is reached with probability 1, i.e., the set of recurrent classes are reached with probability 1.

2. *Property 2.* If $\overline{s}$ is recurrent and it belongs to a recurrent class $\overline{C}$, then $\mathsf{Rec}(\overline{G})(\overline{s}) = \{\overline{C}\}$.

3. *Property 3.* For a recurrent class $\overline{C}$, for all states $\overline{s} \in \overline{C}$, if the Markov chain starts at $\overline{s}$, then all states $\overline{t} \in \overline{C}$ are visited infinitely often with probability 1.

4. *Property 4.* If $\overline{s}'$ is reachable from $\overline{s}$, then $\mathsf{Rec}(\overline{G})(\overline{s}') \subseteq \mathsf{Rec}(\overline{G})(\overline{s})$.

5. *Property 5.* For all $\overline{s}$ we have $\mathsf{Rec}(\overline{G})(\overline{s}) = \bigcup_{(\overline{s},\overline{s}') \in E} \mathsf{Rec}(\overline{G})(\overline{s}')$.

The following lemma is an easy consequence of the above properties.

Lemma 1. *Given a Markov chain* $\overline{G} = (\overline{S}, \overline{\delta})$ *with Muller objective* $\mathsf{Muller}(\mathcal{F})$ *(or a parity objective* $\mathsf{Parity}(p)$*), a state* $\overline{s}$ *is almost-sure winning (resp. positive winning) if for all recurrent classes* $\overline{C} \in \mathsf{Rec}(\overline{G})(\overline{s})$ *(resp. for some recurrent class* $\overline{C} \in \mathsf{Rec}(\overline{G})(\overline{s})$*) reachable from* $\overline{s}$ *we have* $\mathsf{col}(\overline{C}) \in \mathcal{F}$ *(*$\min(p(\overline{C}))$ *is even for the parity objective).*

Proof. From $\overline{s}$ the set of recurrent classes reachable from $\overline{s}$ is reached with probability 1 (Property 1 (b)), and every recurrent class reachable is reached with positive probability. In every recurrent class $\overline{C}$ the minimum priority visited infinitely often with probability 1 is the minimum priority of $\overline{C}$ (Property 3). Also in every recurrent class $\overline{C}$ the set of colors visited infinitely often with probability 1 is exactly the set $\mathsf{col}(\overline{C})$ (Property 3). The desired result follows. $\square$

Markov chains under finite memory strategies. We now define Markov chains obtained by fixing a finite-memory strategy in a POMDP G. A finite-memory strategy $\sigma = (\sigma_u, \sigma_n, M, m_0)$ induces a Markov chain $(S \times M, \delta_\sigma)$, denoted $G \upharpoonright \sigma$, with the probabilistic transition function $\delta_\sigma : S \times M \to \mathcal{D}(S \times M)$: given $s, s' \in S$ and $m, m' \in M$, the transition $\delta_\sigma\big((s', m') \mid (s, m)\big)$ is the probability to go from state (s, m) to state (s', m') in one step under the strategy σ. The probability of transition can be decomposed as follows:

- First an action $a \in A$ is sampled according to the distribution $\sigma_n(m)$;

- then the next state s' is sampled according to the distribution $\delta(s, a)$; and

- finally the new memory m' is sampled according to the distribution $\sigma_u(m, \gamma(s'), a)$ (i.e., the new memory is sampled according σ_u given the old memory, new observation and the action).

More formally, we have:

$$\delta_\sigma\big((s', m') \mid (s, m)\big) = \sum_{a \in A} \sigma_n(m)(a) \cdot \delta(s, a)(s') \cdot \sigma_u(m, \gamma(s'), a)(m').$$

Given $s \in S$ and $m \in M$, we write $(G \upharpoonright \sigma)_{(s,m)}$ for the finite state Markov chain induced on $S \times M$ by the transition function δ_σ, given the initial state is (s, m).

3.2 Belief-based strategies are not sufficient

For all previous decidability results for almost-sure winning in POMDPs, the key was to show that *belief-based* randomized strategies are sufficient. A strategy is belief-based if its memory relies only on the subset construction where the subset denotes the possible current states, i.e., the strategy plays only depending on the set of possible current states of the POMDP, which is called *belief*. In POMDPs with Büchi objectives, belief-based randomized strategies are sufficient for almost-sure winning. We now show with an example that there exist POMDPs with coBüchi objectives, where finite-memory randomized almost-sure winning strategies exist, but there exists no belief-based randomized almost-sure winning strategy.
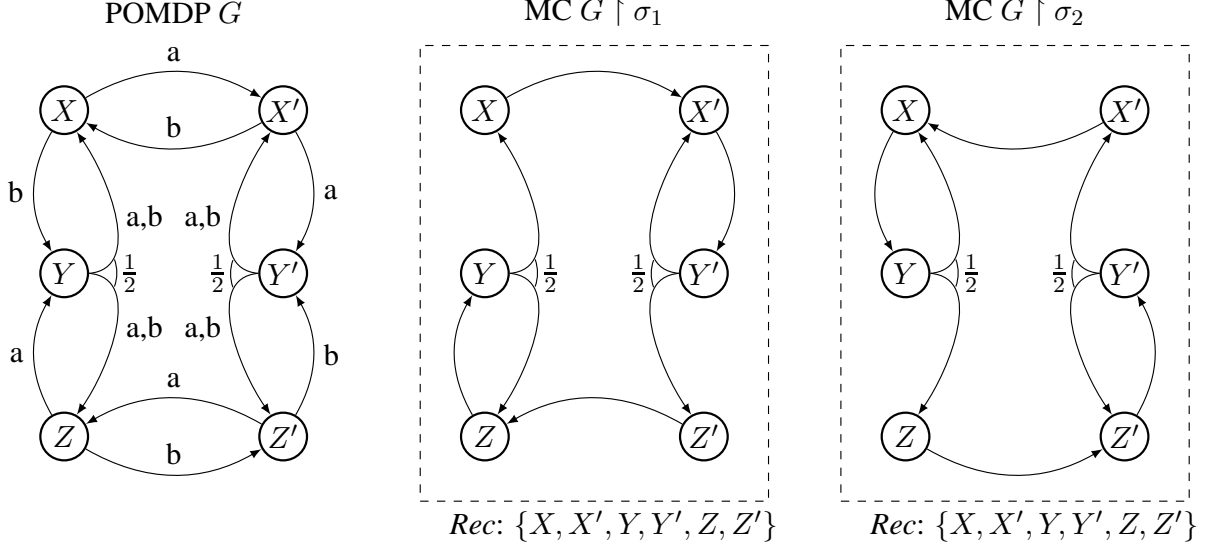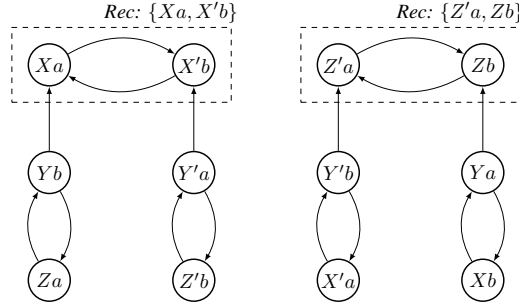
Figure 1: Belief is not sufficient

Figure 2: The Markov chain $G \restriction \sigma_4$.

Example 1. *We consider a POMDP with state space $\{s_0, X, X', Y, Y', Z, Z'\}$ and action set $\{a, b\}$, and let $U = \{X, X', Y, Y', Z, Z'\}$. From the initial state s_0 all the other states are reached with uniform probability in one-step, i.e., for all $s' \in U = \{X, X', Y, Y', Z, Z'\}$ we have $\delta(s_0, a)(s') = \delta(s_0, b)(s') = \frac{1}{6}$. The transitions from the other states are as follows (shown in Figure 1): (i) $\delta(X, a)(X') = 1$ and $\delta(X, b)(Y) = 1$; (ii) $\delta(X', a)(Y') = 1$ and $\delta(X', b)(X) = 1$; (iii) $\delta(Z, a)(Y) = 1$ and $\delta(Z, b)(Z') = 1$; (iv) $\delta(Z', a)(Z) = 1$ and $\delta(Z', b)(Y') = 1$; (v) $\delta(Y, a)(X) = \delta(Y, b)(X) = \delta(Y, a)(Z) = \delta(Y, b)(Z) = \frac{1}{2}$; and (vi) $\delta(Y', a)(X') = \delta(Y', b)(X') = \delta(Y', a)(Z') = \delta(Y', b)(Z') = \frac{1}{2}$. All states in U have the same observation. The coBüchi objective is given by the target set $\{X, X', Z, Z'\}$, i.e., Y and Y' must be visited only finitely often.*

The belief initially after one-step is the set $U = \{X, X', Y, Y', Z, Z'\}$ since from s_0 all of them are reached with positive probability. The belief is always the set U since every state has an input edge for every action, i.e., if the current belief is U (i.e., the set of states that the POMDP is currently in with positive probability is U), then irrespective of whether a or b is chosen all states of U are reached with positive probability and hence the belief set is again U. There are three belief-based strategies: (i) σ_1 that plays
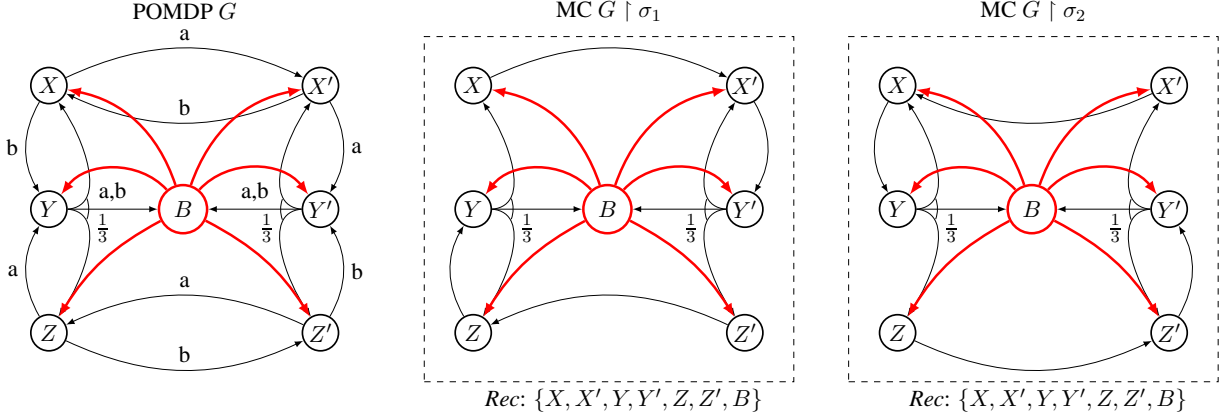
Figure 3: Belief is not sufficient

always a; (ii) σ_2 that plays always b; or (iii) σ_3 that plays both a and b with positive probability. The Markov chains $G \upharpoonright \sigma_1$ (resp. σ_2 and σ_3) are obtained by retaining the edges labeled by action a (resp. action b, and both actions a and b). For all the three strategies, the Markov chains obtained have the whole set U as the recurrent class, and hence both Y and Y' are visited infinitely often with probability 1 violating the coBüchi objective. The Markov chains $G \upharpoonright \sigma_1$ and $G \upharpoonright \sigma_2$ are also shown in Figure 1, and the graph of $G \upharpoonright \sigma_3$ is the same as the POMDP (with edge labels removed). The strategy σ_4 that plays action a and b alternately gives rise to the Markov chain $G \upharpoonright \sigma_4$ (shown in Figure 2) where the recurrent classes do not intersect with Y or Y', and is a finite-memory almost-sure winning strategy. $\square$

In Example 1 the coBüchi objective is not a visible objective. In the following example we modify Example 1 to show that randomized belief-based strategies are not sufficient even if we consider visible coBüchi objectives.

Example 2. *We consider the POMDP shown in Figure 3: the transition edges in the set $U = \{X, X', Y, Y', Z, Z'\}$ are exactly the same as in Figure 1, and the transition probabilities are always uniform over the support set. We add a new state B and from the state Y and Y' add positive transition probabilities (probability $\frac{1}{3}$) to the state B for both actions a and b. Recall that Y and Y' were the bad states in Example 1. From state B all states in U are reached with positive probability for both actions a and b. All states in U have the same observation (denoted as o_U), and the state B has a new and different observation (denoted as o_B). The coBüchi objective is to visit only states with observation o_U infinitely often (i.e., to avoid to visit state B infinitely often). Note that the objective is a visible objective. Since we retain all edges as in Figure 1 and from B all states in U are reached with positive probability in one step, whenever the current observation is o_U, then the belief is the set U. As in Example 1 there are three belief-based strategies (σ_1, σ_2 and σ_3) in belief U, and the Markov chains under σ_1 and σ_2 are shown in Figure 3, and the Markov chain under σ_3 has the same edges as the original POMDP. For all the belief-based strategies the recurrent class contains the state B, and hence B is visited infinitely often with probability 1 violating the coBüchi objective. The strategy σ_4 that alternates actions a and b is a finite-memory almost-sure winning strategy and the Markov chain obtained given σ_4 is shown in Figure 4. Also note that our example shows that belief-based strategies are also not sufficient for positive winning for coBüchi objectives.*
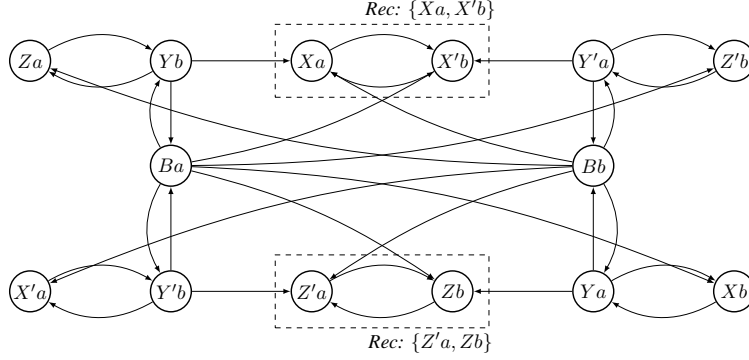
Figure 4: The Markov chain $G \upharpoonright \sigma_4$.

Remark 2. *In Example 2 we have shown that randomized belief-based strategies are not sufficient for finite-memory almost-sure and positive winning strategies in POMDPs with coBüchi objectives. In contrast, for almost-sure winning for Büchi objectives in POMDPs, randomized belief-based strategies are sufficient [9] (both for finite-memory and infinite-memory strategies). The fact that randomized belief-based strategies are not sufficient for finite-memory positive winning strategies in POMDPs with Büchi objectives can be obtained from a simple modification of Example 2 as follows: we consider the POMDP in Example 2 and change the state B to an absorbing state. The Büchi objective is to visit the observation o_U infinitely often, and for all the three belief-based strategies σ_1, σ_2, and σ_3 the Markov chain has only one recurrent class consisting of the absorbing state B. The strategy σ_4 ensures that with positive probability a recurrent class is contained in o_U and is a finite-memory positive winning strategy. Finally, for positive winning in POMDPs with coBüchi objectives, the EXPTIME-complete computational complexity result was obtained with the following straight forward observation [8]: given a POMDP G with a coBüchi objective $\mathsf{coBuchi}(\mathcal{T})$, let S_W be the set of states s such that if s is the starting state (i.e., initial belief is $\{s\}$), then almost-sure safety can be ensured for the target set (i.e., $\mathsf{Safe}(\mathcal{T})$ can be ensured almost-surely). Then positive winning for coBüchi coincides with positive reachability to the set S_W because as soon as S_W is reached, then the current belief contains a state in S_W, and then with positive probability the strategy can assume that the current state is a state in S_W and play the almost-sure safety strategy and the strategy ensures that the coBüchi objective is satisfied with positive probability. Conversely it was also shown that a positive winning strategy for the coBüchi objective must ensure positive probability reachability to S_W [8]. Hence positive winning for coBüchi objectives can be ensured by solving almost-sure safety and positive reachability, and thus we obtain the EXPTIME-complete result from results of almost-sure safety and positive reachability. However, from the previous construction it was not clear whether belief-based strategies are sufficient or not, and Example 2 shows that belief-based strategies are not sufficient.*

3.3 Upper bound on memory of finite-memory strategies

For the following of the section, we fix a POMDP $G = (S, A, \delta, \mathcal{O}, \gamma, s_0)$, with a Muller objective $\mathsf{Muller}(\mathcal{F})$ with the set D of colors and a color mapping function col. We will denote by $\mathfrak{D}$ the powerset of the powerset of the set D of colors, i.e., $\mathfrak{D} = \mathcal{P}(\mathcal{P}(D))$; and note that $|\mathfrak{D}| = 2^{2^d}$, where $d = |D|$. The goal of the section is to prove the following fact: given a finite-memory almost-sure (resp. positive) winning strategy σ on G there exists a finite-memory almost-sure (resp. positive) winning strategy σ' on G, of memory size at most $\mathsf{Mem}^* = 2^{|S|} \cdot 2^{|S|} \cdot |\mathfrak{D}|^{|S|}$.

Overview of the proof. We first present an overview of our proof structure.

- Given an arbitrary finite-memory strategy σ we will consider the Markov chain $G \upharpoonright \sigma$ arising by fixing the strategy.

- Given the Markov chain we will define a projection graph that depends on the recurrent classes of the Markov chain. The projection graph is of size at most Mem^*.

- Given the projection graph we will construct a projected strategy with memory size at most Mem^* that preserves the recurrent classes of the Markov chain $G \upharpoonright \sigma$.

Notations. Given $Z \in \mathfrak{D}^{|S|}$ and given $s \in S$, we write $Z(s)$ (which is in $\mathfrak{D} = \mathcal{P}(\mathcal{P}(D))$) for the s-component of Z. For two sets U_1 and U_2 and $U \subseteq U_1 \times U_2$, we denote by $\mathsf{Proj}_1(U)$ the projection of U on the first component, formally, $\mathsf{Proj}_1(U) = \{u_1 \in U_1 \mid \exists u_2 \in U_2. (u_1, u_2) \in U\}$; and the definition of $\mathsf{Proj}_2(U)$ for the projection on the second component is analogous.

Basic definitions for the projection graph. We now introduce notions associated with the finite Markov chain $G \upharpoonright \sigma$ that will be essential in defining the projection graph.

Definition 3 (Recurrence set functions). *Let σ be a finite-memory strategy with memory M on G for the Muller objective with the set D of colors, and let $m \in M$.*

- *(Function set recurrence). The function $\mathsf{SetRec}_\sigma(m) : S \to \mathfrak{D}$ maps every state $s \in S$ to the projections of colors of recurrent classes reachable from (s, m) in $G \upharpoonright \sigma$. Formally, $\mathsf{SetRec}_\sigma(m)(s) = \{\mathsf{col}(\mathsf{Proj}_1(U)) \mid U \in \mathsf{Rec}(G \upharpoonright \sigma)((s, m))\}$, i.e., we consider the set $\mathsf{Rec}(G \upharpoonright \sigma)((s, m))$ of recurrent classes reachable from the state (s, m) in $G \upharpoonright \sigma$, obtain the projections on the state space S and consider the colors of states in the projected set. We will in sequel consider $\mathsf{SetRec}_\sigma(m) \in \mathfrak{D}^{|S|}$.*

- *(Function boolean recurrence). The function $\mathsf{BoolRec}_\sigma(m) : S \to \{0, 1\}$ is such that for all $s \in S$, we have $\mathsf{BoolRec}_\sigma(m)(s) = 1$ if there exists $U \in \mathsf{Rec}(G \upharpoonright \sigma)((s, m))$ such that $(s, m) \in U$, and 0 if not. Intuitively, $\mathsf{BoolRec}_\sigma(m)(s) = 1$ if (s, m) belongs to a recurrent class in $G \upharpoonright \sigma$ and 0 otherwise. In sequel we will consider $\mathsf{BoolRec}_\sigma(m) \in \{0, 1\}^{|S|}$.*

We first define the projection graph and then present a simple property of $\mathsf{SetRec}_\sigma(m)$ function related to the reachability property.

Definition 4 (Projection graph). *Let σ be a finite-memory strategy. We define the* projection graph *$\mathsf{PrGr}(\sigma) = (V, E)$ associated to σ as follows:*

- *(Vertex set). The set of vertices is $V = \{(U, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m)) \mid U \subseteq S \text{ and } m \in M\}$. Note that $V \subseteq \mathcal{P}(S) \times \{0, 1\}^{|S|} \times \mathfrak{D}^{|S|}$, and hence $|V| \leq \mathsf{Mem}^*$.*

- *(Edge labels). The edges of the graph are labeled by actions in A.*

- *(Edge set). Let $U \subseteq S$, $m \in M$ and $a \in \mathsf{Supp}(\sigma_n(m))$. Let $\overline{U} = \bigcup_{s \in U} \mathsf{Supp}(\delta(s, a))$ denote the set of possible successors of states in U given action a. We add the following set of edges in E: Given (U', m') such that there exists $o \in \mathcal{O}$ with $\gamma^{-1}(o) \cap \overline{U} = U'$ and $m' \in \mathsf{Supp}(\sigma_u(m, o, a))$, we add the edge $(U, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m)) \xrightarrow{a} (U', \mathsf{BoolRec}_\sigma(m'), \mathsf{SetRec}_\sigma(m'))$ to E. Intuitively, the update from U to U' is the update of the belief, i.e., if the previous belief is the set U of states, and the current observation is o, then the new belief is U'; the update of m to m' is according to the support of the memory update function; and the $\mathsf{BoolRec}$ and SetRec functions for the memories are given by the strategy σ.*

- (Initial vertex). *The* initial vertex *of* $\mathsf{PrGr}(\sigma)$ *is the vertex* $(\{s_0\}, \mathsf{BoolRec}_\sigma(m_0), \mathsf{SetRec}_\sigma(m_0))$.

For the rest of this section we fix an arbitrary finite-memory strategy σ that uses memory M.

Lemma 2. *Let* $s, s' \in S$ *and* $m, m' \in M$ *be such that* (s', m') *is reachable from* (s, m) *in* $G \upharpoonright \sigma$. *Then* $\mathsf{SetRec}_\sigma(m')(s') \subseteq \mathsf{SetRec}_\sigma(m)(s)$.

Proof. Since (s', m') is reachable from (s, m) in $G \upharpoonright \sigma$, it follows by simple properties of Markov chains and recurrent classes that we have $\mathsf{Rec}(G \upharpoonright \sigma)((s', m')) \subseteq \mathsf{Rec}(G \upharpoonright \sigma)((s, m))$ (Property 4 of Markov chains). The $\subseteq$ relation is preserved under the projections on states, and then considering the color mapping. Hence the result follows. $\qquad\square$

In the following two lemmas we establish the connection of functions $\mathsf{BoolRec}_\sigma(m)$ and $\mathsf{SetRec}_\sigma(m)$ with the edges of the projection graph. The intuitive description of the first lemma is as follows: it shows that if $\mathsf{BoolRec}$ is set to 1 for a vertex of the projection graph, then for all successors according to the edges of the projection graph, $\mathsf{BoolRec}$ is also 1 for the successors. The second lemma shows a similar result for the projection graph showing that the SetRec functions are subsets for each component for successor vertices.

Lemma 3. *Let* $(V, E) = \mathsf{PrGr}(\sigma)$ *be the projection graph of* σ. *Let* $(U, B, L) \xrightarrow{a} (U', B', L')$ *be an edge in* E, *where* $U, U' \subseteq S$, $B, B' \in \{0,1\}^{|S|}$, *and* $L, L' \in \mathfrak{D}^{|S|}$. *Then for all* $s \in U$ *and* $s' \in \mathrm{Supp}(\delta(s, a))$ *the following assertion holds: if* $B(s) = 1$, *then* $B'(s') = 1$.

Proof. We first note that if $(U, B, L) \xrightarrow{a} (U', B', L')$ is an edge in E, then there exists memory state m and m' such that (i) $B = \mathsf{BoolRec}_\sigma(m)$ and $L = \mathsf{SetRec}_\sigma(m)$; (ii) $B' = \mathsf{BoolRec}_\sigma(m')$ and $L' = \mathsf{SetRec}_\sigma(m')$; (iii) $a \in \mathrm{Supp}(\sigma_n(m))$ and $m' \in \mathrm{Supp}(\sigma_u(m, \gamma(s'), a))$. Hence it follows that (s', m') is reachable in one step from (s, m) in $G \upharpoonright \sigma$. Now, if (s', m') is reached with positive probability from (s, m) in $G \upharpoonright \sigma$ and if (s, m) is a recurrent state of $G \upharpoonright \sigma$, then (s', m') is also recurrent and lies in the same recurrent class as (s, m) (since both (s, m) and (s', m') would lie in the same bottom scc of the graph of the Markov chain). Thus if $\mathsf{BoolRec}_\sigma(m)(s) = 1$, then $\mathsf{BoolRec}_\sigma(m')(s') = 1$. Since $B = \mathsf{BoolRec}_\sigma(m)$ and $B' = \mathsf{BoolRec}_\sigma(m')$, the desired result follows. $\qquad\square$

Lemma 4. *Let* $(V, E) = \mathsf{PrGr}(\sigma)$ *be the projection graph of* σ. *Let* $(U, B, L) \xrightarrow{a} (U', B', L')$ *be an edge in* E, *where* $U, U' \subseteq S$, $B, B' \in \{0,1\}^{|S|}$, *and* $L, L' \in \mathfrak{D}^{|S|}$. *Then for all* $s \in S$ *and all* $s' \in \mathrm{Supp}(\delta(s, a))$, *we have* $L'(s') \subseteq L(s)$.

Proof. By definition of $\mathsf{PrGr}(\sigma)$ (as in the proof of Lemma 3), there exist $m, m' \in M$ such that (i) $B = \mathsf{BoolRec}_\sigma(m)$ and $L = \mathsf{SetRec}_\sigma(m)$; (ii) $B' = \mathsf{BoolRec}_\sigma(m')$ and $L' = \mathsf{SetRec}_\sigma(m')$. Moreover, $a \in \mathrm{Supp}(\sigma_n(m))$, and (U', m') is such that there exists $o \in \mathcal{O}$ with $U' = (\bigcup_{s \in U} \mathrm{Supp}(\delta(s, a)) \cap \gamma^{-1}(o)$ and $m' \in \mathrm{Supp}(\sigma_u(m, o, a))$. This implies that (s', m') is reachable from (s, m) in $G \upharpoonright \sigma$. As a consequence $\mathsf{Rec}(G \upharpoonright \sigma)((s', m')) \subseteq \mathsf{Rec}(G \upharpoonright \sigma)((s, m))$ (Property 4), and as $\subseteq$ relation is preserved by the projection on the states and then on the colors, it follows that $L'(s') \subseteq L(s)$. $\qquad\square$

We now define the notion of projected strategies: intuitively the projected strategy collapses memory with same $\mathsf{BoolRec}$ and SetRec functions, and at a collapsed memory state plays uniformly the union of the actions played at the corresponding memory states.

Definition 5 (Projected strategy $proj(\sigma)$ of a finite-memory strategy). *Let* $\mathsf{PrGr}(\sigma) = (V, E)$ *be the projection graph of* σ. *We define the following projected strategy* $\sigma' = proj(\sigma) = (\sigma_u', \sigma_n', M', m_0')$:

- (Memory set). *The memory set of $proj(\sigma)$ is $M' = V = \{(U, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m)) \mid U \subseteq S$ and $m \in M\}$.*

- (Initial memory). *The initial memory state of $proj(\sigma)$ is $m'_0 = (\{s_0\}, \mathsf{BoolRec}_\sigma(m_0), \mathsf{SetRec}_\sigma(m_0))$.*

- (Memory update). *Let $m = (U, B, L) \in M'$, $o \in \mathcal{O}$ and $a \in A$. Then $\sigma'_u(m, o, a)$ is the uniform distribution over the set $\{m' = (U', B', L') \in M' \mid m \xrightarrow{a} m' \in E$ and $U' \subseteq \gamma^{-1}(o)\}$.*

- (Action selection). *Given $m \in M'$, the action selection function $\sigma'_n(m)$ is the uniform distribution over $\{a \in A \mid \exists m' \in M' \ s.t. \ m \xrightarrow{a} m' \in E\}$.*

Markov chain of the projected strategy. For the following of the section, we fix a finite-memory strategy σ on G, let $(V, E) = \mathsf{PrGr}(\sigma)$ be the projection graph, and let $\sigma' = proj(\sigma)$ be the projected strategy. The finite-memory strategy $\sigma' = (\sigma'_u, \sigma'_n, M', m'_0)$ induces a probability transition function on $S \times M'$: given $s, s' \in S$ and $m, m' \in M'$ let $\delta_{\sigma'}\big((s', m') \mid (s, m)\big)$ be the probability to go from state (s, m) to state (s', m') in one step if we use strategy σ'. Formally,

$$\delta_{\sigma'}\big((s', m') \mid (s, m)\big) = \sum_{a \in A} \sigma'_n(m)(a) \cdot \delta(s, a)(s') \cdot \sigma'_u(m, \gamma(s'), a)(m').$$

The chain $G \restriction \sigma'$ is a finite state Markov chain, with state space $S \times M'$, which is a subset of $S \times \mathcal{P}(S) \times \{0, 1\}^{|S|} \times \mathfrak{D}^{|S|}$. Given $X \in S$, $Y \subseteq S$, $C \in \{0, 1\}^{|S|}$, and $Z \in \mathfrak{D}^{|S|}$, let $\mathsf{Succ}_1((X, Y, C, Z))$ denote the set of states of the Markov chain reachable in one step from the state (X, Y, C, Z).

Random variable notations. For all $n \geq 0$ we write X_n, Y_n, C_n, Z_n, W_n for the random variables which correspond respectively to the projection of the n-th state of the Markov chain $G \restriction \sigma'$ on the S component, the $\mathcal{P}(S)$ component, the $\{0, 1\}^{|S|}$ component, the $\mathfrak{D}^{|S|}$ component, and the n-th action, respectively.

Run of the Markov chain $G \restriction \sigma'$. A *run* on $G \restriction \sigma'$ is a sequence

$$r = (X_0, Y_0, C_0, Z_0) \xrightarrow{W_0} (X_1, Y_1, C_1, Z_1) \xrightarrow{W_1} \ldots$$

such that each finite prefix of r is generated with positive probability on the chain, i.e., for all $i \geq 0$, we have (i) $W_i \in \mathrm{Supp}(\sigma'_n(Y_i, C_i, Z_i))$; (ii) $X_{i+1} \in \mathrm{Supp}(\delta(X_i, W_i))$; and (iii) $(Y_{i+1}, C_{i+1}, Z_{i+1}) \in \mathrm{Supp}(\sigma'_u((Y_i, C_i, Z_i), \gamma(X_{i+1}), W_i))$. In the following three lemmas we establish crucial properties of the Markov chain obtained from the projected strategy.

Lemma 5. *Let $X \in S$, $Y \subseteq S$, $C \in \{0, 1\}^{|S|}$, and $Z \in \mathfrak{D}^{|S|}$. Then:*

$$Z(X) = \bigcup_{(X', Y', C', Z') \in \mathsf{Succ}_1((X, Y, C, Z))} Z'(X').$$

Proof. This follows from the following basic property of finite Markov chains: given a state s of a finite Markov chain, the set of recurrent classes reachable from s is the union of the set of recurrent classes reachable from the set of states reachable from s in one step (Property 5 of Markov chains). The relation is preserved by projection on the colors of states. $\quad\square$

Lemma 6. *Let $\sigma' = proj(\sigma)$ be the projected strategy of σ. Given $s, s' \in S$ and $m, m' \in M$, if (s', m') is reachable from (s, m) in $G \restriction \sigma$, then for all $Y \subseteq S$ such that $(s, Y, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m))$ is a state of $G \restriction \sigma'$, there exists $Y' \subseteq S$ such that $(s', Y', \mathsf{BoolRec}_\sigma(m'), \mathsf{SetRec}_\sigma(m'))$ is reachable from $(s, Y, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m))$ in $G \restriction \sigma'$.*

Proof. Suppose first that (s', m') is reachable from (s, m) in $G \upharpoonright \sigma$ in one step. Let $Y \subseteq S$ be such that $(s, Y, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m))$ is a state of $G \upharpoonright \sigma'$. Then there exists an edge in the projection graph of σ from $(Y, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m))$ to another vertex $(Y', \mathsf{BoolRec}_\sigma(m'), \mathsf{SetRec}_\sigma(m'))$. As a consequence, there exists $Y' \subseteq S$ such that $(s', Y', \mathsf{BoolRec}_\sigma(m'), \mathsf{SetRec}_\sigma(m'))$ is reachable from $(s, Y, \mathsf{BoolRec}_\sigma(m), \mathsf{SetRec}_\sigma(m))$ in $G \upharpoonright \sigma'$.

We conclude the proof by induction: if (s', m') is reachable from (s, m) in $G \upharpoonright \sigma$, then there exists a sequence of couples $(s_1, m_1), (s_2, m_2), ..., (s_i, m_i)$ such that $(s_1, m_1) = (s, m)$, $(s_i, m_i) = (s', m')$, and for all $j \in \{1, ..., i-1\}$ we have that (s_{j+1}, m_{j+1}) is reachable from (s_j, m_j) in one step. Using the proof for an elementary step (or one step) inductively on such a sequence, we get the result. $\square$

Lemma 7. *Let $X_0 \in S$, $Y_0 \in \mathcal{P}(S)$, $C_0 \in \{0,1\}^{|S|}$ and $Z_0 \in \mathfrak{D}^{|S|}$, and let $r = (X_0, Y_0, C_0, Z_0) \overset{W_0}{\to} (X_1, Y_1, C_1, Z_1) \overset{W_1}{\to} ...$ be a run on $G \upharpoonright \sigma'$ with a starting state (X_0, Y_0, C_0, Z_0). Then for all $n \geq 0$ the following assertions hold:*

1. *$X_{n+1} \in \mathrm{Supp}(\delta(X_n, W_n))$.*

2. *$Z_n(X_n)$ is not empty.*

3. *$Z_{n+1}(X_{n+1}) \subseteq Z_n(X_n)$.*

4. *$(Y_n, C_n, Z_n) \overset{W_n}{\to} (Y_{n+1}, C_{n+1}, Z_{n+1})$ is an edge in E, where $(V, E) = \mathsf{PrGr}(\sigma)$.*

5. *If $C_n(X_n) = 1$, then $C_{n+1}(X_{n+1}) = 1$.*

6. *If $C_n(X_n) = 1$, then $|Z_n(X_n)| = 1$. If $\{Z\} = Z_n(X_n)$, then for all $j \geq 0$ we have $\mathsf{col}(X_{n+j}) \in Z$.*

Proof. We prove all the points below.

1. The first point is a direct consequence of the definition of the Markov chain.

2. The second point follows also from the definition of the chain as from every state of a Markov chain at least one recurrent class is reachable and hence the projection on colors is not empty.

3. The third point follows from the first point of the lemma and Lemma 4.

4. For the fourth point: given (X_n, Y_n, C_n, Z_n), the strategy σ'_n samples W_n with uniform probability among the elements of the set:

$$\{a \in A \mid \exists m \in M \text{ s.t. } (Y_n, C_n, Z_n) \overset{a}{\to} m \in E\}.$$

Once W_n has been chosen, then $\sigma'_u((Y_n, C_n, Z_n), \gamma(X_{n+1}), W_n)$ samples $(Y_{n+1}, C_{n+1}, Z_{n+1})$ uniformly among the elements of the set:

$$\{(U, B, L) \in \mathcal{P}(S) \times \{0,1\}^{|S|} \times \mathfrak{D}^{|S|} \mid (Y_n, C_n, Z_n) \overset{W_n}{\to} (U, B, L) \in E \text{ and } U \subseteq \gamma^{-1}(\gamma(X_{n+1}))\}.$$

This proves that $(Y_n, C_n, Z_n) \overset{W_n}{\to} (Y_{n+1}, C_{n+1}, Z_{n+1})$ is an edge in E.

5. The fifth point follows from the fourth point and Lemma 3.

6. Suppose (X_n, Y_n, C_n, Z_n) is such that $C_n(X_n) = 1$. Let $m \in M$ be an arbitrary memory state such that $C_n = \mathsf{BoolRec}_\sigma(m)$ and $Z_n = \mathsf{SetRec}_\sigma(m)$. By hypothesis, since $C_n(X_n) = 1$, it follows that (X_n, m) is a recurrent state in the Markov chain $G \upharpoonright \sigma$. As a consequence, only one recurrent class $R \subseteq S \times M$ of $G \upharpoonright \sigma$ is reachable from (X_n, m), and (X_n, m) belongs to this class (Property 2 of Markov chains). Hence $Z_n(X_n) = \{\mathsf{col}(\mathsf{Proj}_1(R))\}$, and thus $|Z_n(X_n)| = 1$. It also follows that all states (X', m') reachable in one step from (X_n, m) also belong to the recurrent class R. It follows that $X_{n+1} \in \mathsf{Proj}_1(R)$ and hence $\mathsf{col}(X_{n+1}) \in \mathsf{col}(\mathsf{Proj}_1(R))$. By induction for all $j \geq 0$ we have $\mathsf{col}(X_{n+j}) \in \mathsf{col}(\mathsf{Proj}_1(R))$.

The desired result follows. $\qquad\square$

We now introduce the final notion that is required to complete the proof. The notion is that of a pseudo-recurrent state. Intuitively a state (X, Y, C, Z) is pseudo-recurrent if Z contains exactly one recurrent subset, X belongs to the subset and it will follow for some memory $m \in M$ (of certain desired property) (X, m) is a recurrent state in the Markov chain $G \upharpoonright \sigma$. The important property that will be useful is that once a pseudo-recurrent state is reached, then C and Z remain invariant. We now formally define pseudo-recurrent states.

Definition 6 (Pseudo-recurrent states). *Let $X \in S$, $Y \subseteq S$, $C \in \{0, 1\}^{|S|}$, and $Z \in \mathfrak{D}^{|S|}$. Then the state (X, Y, C, Z) is called* pseudo-recurrent *if there exists $Z_\infty \subseteq D$ such that:*

$$\text{(i) } Z(X) = \{Z_\infty\}, \text{ (ii) } \mathsf{col}(X) \in Z_\infty, \text{ and (iii) } C(X) = 1.$$

The following lemma shows that in the Markov chain $G \upharpoonright \sigma'$, all states reachable from a pseudo-recurrent state are also pseudo-recurrent.

Lemma 8. *Let (X, Y, C, Z) be a pseudo-recurrent state. If (X', Y', C', Z') is reachable from (X, Y, C, Z) in $G \upharpoonright \sigma'$, then (X', Y', C', Z') is also a pseudo-recurrent state and $Z'(X') = Z(X)$.*

Proof. Let (X, Y, C, Z) be a pseudo-recurrent state, and let $Z_\infty \subseteq D$ be such that $Z(X) = \{Z_\infty\}$, $\mathsf{col}(X) \in Z_\infty$, and $C(X) = 1$. By Lemma 7 (fifth point), if $C(X) = 1$, then $C'(X') = 1$. By Lemma 7 (third point) also, $Z'(X') = \{Z_\infty\}$, since $Z'(X')$ is a non empty subset of $Z(X)$. Finally, the fact that $\mathsf{col}(X') \in Z_\infty$ follows from the last (sixth) point of Lemma 7. $\qquad\square$

In the following lemma we show that with probability 1 a pseudo-recurrent state is reached in $G \upharpoonright \sigma'$.

Lemma 9. *Let $X \in S$, $Y \in \mathcal{P}(S)$, $C \in \{0, 1\}^{|S|}$, and $Z \in \mathfrak{D}^{|S|}$. Then almost-surely (with probability 1) a run on $G \upharpoonright \sigma'$ from any starting state (X, Y, C, Z) reaches a pseudo-recurrent state.*

Proof. We show that given (X, Y, C, Z) there exists a pseudo-recurrent state (X', Y', C', Z') which is reachable from (X, Y, C, Z) in $G \upharpoonright \sigma'$. First let us consider the Markov chain $G \upharpoonright \sigma$ obtained from the original finite-memory strategy σ with memory M. Let $m \in M$ be such that $C = \mathsf{BoolRec}_\sigma(m)$ and $Z = \mathsf{SetRec}_\sigma(m)$. We will now show that the result is a consequence of Lemma 6. First we know that there exists $t \in S$ and $m' \in M$ such that (t, m') is recurrent and reachable from (X, m) with positive probability in $G \upharpoonright \sigma$. Let $R \subseteq S \times M$ be the unique recurrent class such that $(t, m') \in R$, and $Z_\infty = \{\mathsf{col}(\mathsf{Proj}_1(R))\}$. By Lemma 6, this implies that from (X, Y, C, Z) we can reach a state (X', Y', C', Z') such that:

- $X' = t$;

- $Z'(X') = \{Z_\infty\}$;

- $\mathsf{col}(X') \in Z_\infty$; and

- $C'(X') = 1$.

Hence (X', Y', C', Z') is a pseudo-recurrent state. This shows that from all states with positive probability a pseudo-recurrent state is reached, and since it holds for all states with positive probability, it follows that it holds for all states with probability 1 (Property 1 (a)). □

In the following three lemmas we establish the required properties of pseudo-recurrent states.

Lemma 10. *Let (X, Y, C, Z) be a state of $G \restriction \sigma'$, and let $Z_B \in Z(X)$. Then there exists a pseudo-recurrent state (X', Y', C', Z') which is reachable from (X, Y, C, Z) and such that $Z'(X') = \{Z_B\}$.*

Proof. The proof is of similar flavor as Lemma 9. Consider the Markov chain $G \restriction \sigma$ arising by fixing the original finite-memory strategy σ. Let $m \in M$ such that $C = \mathsf{BoolRec}_\sigma(m)$ and $Z = \mathsf{SetRec}_\sigma(m)$. We have $Z_B \in \mathsf{SetRec}_\sigma(m)(X)$, hence $Z_B = \mathsf{col}(\mathsf{Proj}_1(R))$ for some recurrent class R of the chain $G \restriction \sigma$ reachable from (X, m). Let $t \in S$ and $m' \in M$ be such that (t, m') is reachable from (X, m) in $G \restriction \sigma$ and $\mathsf{Rec}(G \restriction \sigma)((t, m')) = \{R\}$, then $\mathsf{SetRec}_\sigma(m')(t) = \{Z_B\}$. By Lemma 6, there exists Y', C' such that $(t, Y', C', \mathsf{SetRec}_\sigma(m'))$ is reachable from $(X, Y, C, \mathsf{SetRec}_\sigma(m))$ in $G \restriction \sigma'$ from the starting state (X, Y, C, Z). The desired result follows. □

Lemma 11. *Let (X, Y, C, Z) be a pseudo-recurrent state, and $Z_\infty \subseteq D$ such that $Z(X) = \{Z_\infty\}$. Then for all colors $\ell \in Z_\infty$, there exists a state (X', Y', C', Z') which is reachable in $G \restriction \sigma'$ from (X, Y, C, Z) and such that $\mathsf{col}(X') = \ell$.*

Proof. We again consider the Markov chain $G \restriction \sigma$. Let $m \in M$ be such that $C = \mathsf{BoolRec}_\sigma(m)$ and $Z = \mathsf{SetRec}_\sigma(m)$. Let R be the unique recurrent class in $G \restriction \sigma$ such that $(X, m) \in R$, then $Z_\infty = \mathsf{col}(\mathsf{Proj}_1(R))$. For every $\ell \in Z_\infty$, there exists $(X', m') \in R$ such that $\mathsf{col}(X') = \ell$. As (X', m') is reachable from (X, m) in $G \restriction \sigma$, by Lemma 6, there exists Y', C', Z' such that (X', Y', C', Z') is reachable in $G \restriction \sigma'$ from (X, Y, C, Z). □

Lemma 12. *Let (X, Y, C, Z) be a pseudo-recurrent state, then we have $Z(X) = \mathsf{SetRec}_{\sigma'}(m')(X)$, where $m' = (Y, C, Z)$. In other words, if we consider a pseudo-recurrent state, and then consider the projection on the state space of the POMDP G of the recurrent classes reachable and consider the colors, then they coincide with $Z(X)$.*

Proof. Let (X, Y, C, Z) be a pseudo-recurrent state, and let Z_∞ be such that $Z(X) = \{Z_\infty\}$. First, by Lemma 8, we know that if (X', Y', C', Z') is reachable from (X, Y, C, Z) in $G \restriction \sigma'$, then $\mathsf{col}(X') \in Z_\infty$. This implies that for all $Z_B \in \mathsf{SetRec}_{\sigma'}(m')(X)$, where $m' = (Y, C, Z)$, we have $Z_B \subseteq Z_\infty$. Second, by Lemma 11, if (X', Y', C', Z') is reachable from (X, Y, C, Z) in $G \restriction \sigma'$ and $\ell \in Z_\infty$, then there exists (X'', Y'', C'', Z'') reachable from (X', Y', C', Z') such that $\mathsf{col}(X'') = \ell$. This implies that for all $Z_B \in \mathsf{SetRec}_{\sigma'}(m')(X)$, where $m' = (Y, C, Z)$, we have $Z_\infty \subseteq Z_B$. Thus, $\mathsf{SetRec}_{\sigma'}(m')(X) = \{Z_\infty\} = Z(X)$. □

We are now ready to prove the main lemma which shows that the color sets of the projections of the recurrent classes on the state space of the POMDP coincide for σ and the projected strategy $\sigma' = proj(\sigma)$.

Lemma 13. *Consider a finite-memory strategy* $\sigma = (\sigma_u, \sigma_n, M, m_0)$ *and the projected strategy* $\sigma' = proj(\sigma) = (\sigma'_u, \sigma'_n, M', m'_0)$. *Then we have*

$$\mathsf{SetRec}_{\sigma'}(m'_0)(s_0) = \mathsf{SetRec}_{\sigma}(m_0)(s_0);$$

i.e., the colors of the projections of the recurrent classes of the two strategies on the state space of the POMDP G coincide.

Proof. For the proof, let $X = s_0$, $Y = \{s_0\}$, $C = \mathsf{BoolRec}_{\sigma}(m_0)$, $Z = \mathsf{SetRec}_{\sigma}(m_0)$. We need to show that $\mathsf{SetRec}_{\sigma'}(m'_0)(X) = Z(X)$, where $m'_0 = (Y, C, Z)$. We show inclusion in both directions.

- *First inclusion:* ($Z(X) \subseteq \mathsf{SetRec}_{\sigma'}(m'_0)(X)$). Let $Z_B \in Z(X)$. By Lemma 10, there exists a state (X', Y', C', Z') which is reachable in $G \restriction \sigma'$ from (X, Y, C, Z), which is pseudo-recurrent, and such that $Z'(X') = \{Z_B\}$. By Lemma 12, we have $Z'(X') = \mathsf{SetRec}_{\sigma'}(m')(X')$ where $m' = (Y', C', Z')$. By Lemma 2, we have $\mathsf{SetRec}_{\sigma'}(m')(X') \subseteq \mathsf{SetRec}_{\sigma'}(m'_0)(X)$. This proves that $Z_B \in \mathsf{SetRec}_{\sigma'}(m'_0)(X)$.

- *Second inclusion:* ($\mathsf{SetRec}_{\sigma'}(m'_0)(X) \subseteq Z(X)$). Conversely, let $Z_B \in \mathsf{SetRec}_{\sigma'}(m'_0)(X)$. Since $G \restriction \sigma'$ is a finite Markov chain, there exists a state (X', Y', C', Z') which is reachable from (X, Y, C, Z) in $G \restriction \sigma'$ such that:

  - $\{Z_B\} = \mathsf{SetRec}_{\sigma'}(m')(X')$, where $m' = (Y', C', Z')$.
  - For all (X'', Y'', C'', Z'') reachable from (X', Y', C', Z') in $G \restriction \sigma'$ we have $\{Z_B\} = \mathsf{SetRec}_{\sigma'}(m'')(X'')$ where $m'' = (Y'', C'', Z'')$.

  The above follows from the following property of a finite Markov chain: given a state s of a finite Markov chain and a recurrent class R reachable from s, from all states t of R the recurrent class reachable from t is R only (Property 2 of Markov chains). The condition is preserved by a projection on colors of states in R.

  By Lemma 9, there exists a pseudo-recurrent state (X'', Y'', C'', Z'') which is reachable from (X', Y', C', Z') in $G \restriction \sigma'$. By Lemma 12, we know that $Z''(X'') = \mathsf{SetRec}_{\sigma'}(m'')(X'')$ where $m'' = (Y'', C'', Z'')$. Since $\mathsf{SetRec}_{\sigma'}(m'')(X'') = \{Z_B\}$, and since by Lemma 7 (third point) we have $Z''(X'') \subseteq Z'(X') \subseteq Z(X)$, we get that $Z_B \in Z(X)$.

The desired result follows. $\qed$

Theorem 2. *Given a POMDP G and a Muller objective* $\mathsf{Muller}(\mathcal{F})$ *with the set D of colors, the following assertions hold:*

1. *If there is a finite-memory almost-sure winning strategy* σ, *then the projected strategy* $proj(\sigma)$, *with memory of size at most* $\mathsf{Mem}^* = 2^{2 \cdot |S|} \cdot |\mathfrak{D}|^{|S|}$ *(where* $\mathfrak{D} = \mathcal{P}(\mathcal{P}(D))$), *is also an almost-sure winning strategy.*

2. *If there is a finite-memory positive winning strategy* σ, *then the projected strategy* $proj(\sigma)$, *with memory of size at most* Mem^*, *is also a positive winning strategy.*

3. *Finite-memory almost-sure (resp. positive) winning strategies require at least exponential memory in general, and randomized belief-based strategies are not sufficient in general for finite-memory almost-sure and positive winning strategies.*

Proof. Consider a finite-memory strategy σ with memory M and initial memory m_0 and the Markov chain $G \upharpoonright \sigma$. Also recall that the number of memory states used by $proj(\sigma)$ is at most Mem*.

1. By Lemma 1 if σ is almost-sure winning, then for all recurrent classes C reachable from (s_0, m_0) in $G \upharpoonright \sigma$ we have col($\mathsf{Proj}_1(C)) \in \mathcal{F}$; and by Lemma 13 it follows that in the Markov chain $G \upharpoonright proj(\sigma)$ for all recurrent classes C' reachable from (s_0, m'_0), where m'_0 is the initial memory of $proj(\sigma)$, we have col($\mathsf{Proj}_1(C')) \in \mathcal{F}$. It follows from Lemma 1 that $proj(\sigma)$ is an almost-sure winning strategy.

2. By Lemma 1 if σ is positive winning, then there exists some recurrent class C reachable from (s_0, m_0) in $G \upharpoonright \sigma$ with col($\mathsf{Proj}_1(C)) \in \mathcal{F}$; and by Lemma 13 it follows that in the Markov chain $G \upharpoonright proj(\sigma)$ there exists some recurrent class C' reachable from (s_0, m'_0), where m'_0 is the initial memory of $proj(\sigma)$, with col($\mathsf{Proj}_1(C')) \in \mathcal{F}$. It follows from Lemma 1 that $proj(\sigma)$ is a positive winning strategy.

3. The exponential memory requirement follows from the results of [8] that shows exponential memory requirement for almost-sure winning strategies for reachability objectives and positive winning strategies for safety objectives. The fact that randomized belief-based strategies are not sufficient follows from Example 1.

The desired result follows. $\qquad\square$

4 Strategy Complexity for Parity Objectives under Finite-memory Strategies

In this section we will establish the exponential upper bounds for almost-sure (resp. positive) winning strategies in POMDPs with parity objectives under finite-memory strategies. Observe that Büchi and coBüchi objectives are parity (hence also Muller) objectives with 2 priorities (or colors) (i.e., $d = 2$), and from Theorem 2 we already obtain an upper bound of $2^{6 \cdot |S|}$ on memory size for them. However, applying the general result of Theorem 2 for Muller objectives to parity objectives we obtain a double exponential bound, and we will establish the exponential memory bound. Formally we will establish Theorem 3: for item (1), in Section 4.1 we present a reduction that for positive winning given a POMDP with $|S|$ states and a parity objective with $2 \cdot d$ priorities constructs an equivalent POMDP with $d \cdot |S|$ states with Büchi objectives (and thus applying Theorem 2 we obtain the $2^{3 \cdot d \cdot |S|}$ upper bound); for item (2), in Section 4.2 we present a reduction that for almost-sure winning given a POMDP with $|S|$ states and a parity objective with $2 \cdot d$ priorities constructs an equivalent POMDP with $d \cdot |S|$ states with a coBüchi objective (and thus applying Theorem 2 we obtain the $2^{3 \cdot d \cdot |S|}$ upper bound); and item (3) follows as in the proof of Theorem 2.

Theorem 3. *Given a POMDP G and a parity objective* Parity(p) *with the set D of d priorities, the following assertions hold:*

1. *If there is a finite-memory positive winning strategy, then there is a positive winning strategy with memory of size at most $2^{3 \cdot d \cdot |S|}$.*

2. *If there is a finite-memory almost-sure winning strategy, then there is an almost-sure winning strategy with memory of size at most $2^{3 \cdot d \cdot |S|}$.*

3. *Finite-memory almost-sure (resp. positive) winning strategies require exponential memory in general, and randomized belief-based strategies are not sufficient in general for finite-memory almost-sure (resp. positive) winning strategies.*

4.1 Positive parity to positive Büchi

Given a POMDP $G = (S, A, \delta, \mathcal{O}, \gamma, s_0)$ and a parity objective $\mathsf{Parity}(p)$ with priority set $D = \{0, \ldots, 2 \cdot d\}$, we will construct a POMDP $G' = (S', A, \delta', \mathcal{O}', \gamma', s_0')$ together with a Büchi objective $\mathsf{Buchi}(\mathcal{T})$ such that positive winning under finite-memory strategies is preserved. Let I denote the set $\{0, \ldots, d\}$. Intuitively, in the construction of G' we form a copy G_i of the POMDP G for each $i \in I$. There will be a positive probability of going from the newly added initial state s_0' to every copy G_i. The transition probabilities in the copy G_i will mimic the transition probability of G for states with priority at least $2 \cdot i$, and for states with priority strictly smaller than $2 \cdot i$ it mimics the transitions of G with probability $1/2$ and with probability $1/2$ goes to a newly added absorbing state s_f. Note that the construction will ensure that for any finite-memory strategy, in the copy G_i there are no recurrent classes that contain priorities strictly smaller than $2 \cdot i$ as the absorbing state s_f is always reached with positive probability from such states (with priority strictly smaller than $2 \cdot i$). Note that every recurrent class that intersects with an absorbing state must consists only of the absorbing state, since there are no transitions from the absorbing state to any other state. In the copy G_i states with priority $2 \cdot i$ are assigned priority 0, and all other states are assigned priority 1. Formally the construction is as follows:

- $S' = (S \times I) \cup \{s_0', s_f\}$.

- We define the probabilistic transition function δ' as follows:

  1. $\delta'(s_0', a)((s, i)) = \frac{\delta(s_0, a)(s)}{|I|}$, for all $a \in A$ and all $i \in I$, i.e., with positive probability we move to copy G_i for all $i \in I$;

  2. $\delta'((s, i), a)((s', i)) = \begin{cases} \delta(s, a)(s') & \text{if } p(s) \geq 2 \cdot i; \\ \frac{\delta(s, a)(s')}{2} & \text{otherwise;} \end{cases}$

  3. if $p(s) < 2 \cdot i$, then we also have $\delta'((s, i), a)(s_f) = \frac{1}{2}$;

  4. $\delta'(s_f, a)(s_f) = 1$ for all $a \in A$ (i.e., s_f is an absorbing state).

- $\mathcal{O}' = \mathcal{O} \cup \{s_f\}$.

- $\gamma'((s, i)) = \gamma(s)$, $\gamma'(s_0') = \gamma(s_0)$ and $\gamma'(s_f) = s_f$.

We define the priority function p' for the Büchi objective as follows:

1. $p'(s_0') = 1$;

2. $p'((s, i)) = \begin{cases} 0 & \text{if } p(s) = 2 \cdot i; \\ 1 & \text{otherwise;} \end{cases}$

3. $p'(s_f) = 1$.

The Büchi objective for G' is $\mathsf{Buchi}(p'^{-1}(0))$, i.e., the target set $\mathcal{T}$ is the set of states with priority 0 according to p'.

Lemma 14. *If there exists a finite-memory positive winning strategy in G for the parity objective $\mathsf{Parity}(p)$, then there exists a finite-memory positive winning strategy with the same memory states in G' for the objective $\mathsf{Buchi}(p'^{-1}(0))$.*

Proof. Let $\sigma = (\sigma_u, \sigma_n, M, m_0)$ be a finite-memory positive winning strategy in the POMDP G for the objective $\mathsf{Parity}(p)$. We define the strategy $\sigma' = (\sigma'_u, \sigma'_n, M, m_0)$ in the POMDP G' as follows: the strategy plays as σ for all states other than s_f, and $\sigma'_u(\{s_f\}, m, a)(m)$ for all $a \in A$. As the only state in the observation s_f is the absorbing state s_f, no matter what the strategy plays, s_f is not left. The rest of the components is the same as in the strategy σ. Let $\widehat{G}$ denote the Markov Chain $G \upharpoonright \sigma$ and $\widehat{G}'$ the Markov chain $G' \upharpoonright \sigma'$.

Reachability. We first show that if (s', m') is reachable from (s, m) in $\widehat{G}$ for $s, s' \in S$ and $m, m' \in M$, then for all $i \in I$ we have that (s', i, m') is reachable from (s, i, m) in $\widehat{G}'$. We prove the fact inductively. Let $(s, m) \rightarrow (s', m')$ be an edge in $\widehat{G}$, then there exists an action $a \in A$ such that (i) $s' \in \mathrm{Supp}(\delta(s, a))$, (ii) $a \in \mathrm{Supp}(\sigma_n(m))$, and (iii) $m' \in \mathrm{Supp}(\sigma_u(m, \gamma(s'), a))$. By definition of the transition function δ' this implies that $(s', i) \in \mathrm{Supp}(\delta'((s, i), a)$. Since σ' plays the same as σ, it follows that $(s, i, m) \rightarrow (s', i, m')$ is an edge in $\widehat{G}'$. As the state spaces of the Markov chains are finite, this extends to reachability by simple induction.

Recurrent class. Since σ is a positive winning strategy, there must exist a recurrent class C reachable from (s_0, m_0) in $\widehat{G}$ such that the minimal priority $\min(p(\mathsf{Proj}_1(C)))$ is even. Let that minimal priority be $2 \cdot i$. Consider the copy G_i of G in G': it contains all states from $\mathsf{Proj}_1(C)$, and moreover as the minimal priority of the states in $\mathsf{Proj}_1(C)$ is $2 \cdot i$ (according to p), the transition function δ' matches the transition function δ for states in $\mathsf{Proj}_1(C)$ and all actions $a \in A$. As the strategy σ' does not know, due to the observation definition, in which copy it is and plays as the strategy σ does in G, the set $C' = \{(s, i, m) \mid (s, m) \in C\}$ of states forms a recurrent class in $\widehat{G}'$.

Finally we need to show that C' is reachable from (s'_0, m_0) in $\widehat{G}'$. Since C is reachable from (s_0, m_0) in $\widehat{G}$, there exists a state (s, m) that is reachable in one step from (s_0, m_0) and C is reachable from (s, m). The state (s, i, m) is reachable in one step from (s'_0, m_0) in $\widehat{G}'$ (from the initial state s'_0 all copies are reached with positive probability), and reachability to C' from (s, i, m) follows from the argument above for reachability. As the set $\mathsf{Proj}_1(C')$ contains a state s with $p(s) = 2 \cdot i$, we have that $p'((s, i)) = 0$, i.e., a target state belongs to C'. It follows that σ' is a finite-memory positive winning strategy in G' for $\mathsf{Buchi}(p'^{-1}(0))$ and the desired result follows. $\qquad\square$

Lemma 15. *If there exists a finite-memory positive winning strategy in G' for the objective $\mathsf{Buchi}(p'^{-1}(0))$, then there exists a finite-memory positive winning strategy with the same memory states in G for the objective $\mathsf{Parity}(p)$.*

Proof. Given a finite-memory positive winning strategy $\sigma = (\sigma_u, \sigma_n, M, m_0)$ in the POMDP G' we show that σ is also positive winning in the POMDP G. Similar to the previous lemma we fix the strategy σ in G and obtain a Markov Chain $\widehat{G} = G \upharpoonright \sigma$ and $\widehat{G}' = G' \upharpoonright \sigma$.

Reachability. We show that if (s', i, m') is reachable from (s, i, m) in $\widehat{G}'$ for $s, s' \in S$, $m, m' \in M$, and $i \in I$, then (s', m') is reachable from (s, m) in $\widehat{G}$. This follows from the fact that (i) if there is an edge $(s, i) \xrightarrow{a} (s', i)$ in $\widehat{G}'$ (i.e., $\delta'((s, i), a)((s', i)) > 0$), then we have an edge $s \xrightarrow{a} s'$ in $\widehat{G}$ (i.e., $\delta(s, a)(s') > 0$); and (ii) the strategy σ is the same for both POMDPs.

Recurrent class. As σ is a positive winning strategy in G', there exists a recurrent class C' reachable from (s'_0, m_0) in $\widehat{G}'$ which satisfies $\mathsf{Proj}_1(C') \cap p'^{-1}(0)$ is non-empty. Note that there must exist an $i \in I$ such that all the states of the recurrent class C' are elements from the set $S \times \{i\} \times M$, i.e., the class is included in some copy G_i (since there are no transitions between copies and the absorbing state has priority 1). As C' forms a recurrent class in copy i it follows that all the states in $\mathsf{Proj}_1(C')$ have priority at least $2 \cdot i$ according to p (since states with priority strictly smaller than $2 \cdot i$ according to p have positive transition probability to s_f).

Consider the set of states $C = \{(s,m) \mid (s,i,m) \in C'\}$ in $\widehat{G}$. As the strategies we consider are the same, C forms a recurrent class in $\widehat{G}$ with the minimal priority at least $2 \cdot i$. Moreover, since $\mathsf{Proj}_1(C') \cap p'^{-1}(0)$ is non-empty, it follows that the minimal priority of C is exactly $2 \cdot i$, i.e., $\min(p(\mathsf{Proj}_1(C)))$ is $2 \cdot i$ and even.

Finally, it remains to show that C is reachable from the initial state of $\widehat{G}$. Since C' is reachable from (s'_0, m_0), it must be reachable from some state (s, i, m) of copy G_i and (s, i, m) is reachable in one step from (s'_0, m_0) in $\widehat{G}'$. Then it follows that the state (s, m) is reachable from (s_0, m_0) in one step in $\widehat{G}$, and the reachability of C from (s, m) follows from the fact that C' is reachable from (s, i, m) and the argument for reachability above. Hence, σ is a positive winning strategy for the objective $\mathsf{Parity}(p)$ in G and the desired result follows. $\qquad\square$

Lemma 14 and Lemma 15 establishes item (1) of Theorem 3.

4.2 Almost-sure parity to almost-sure coBüchi

For almost-sure winning the reduction from parity objectives to coBüchi objectives will be achieved in two steps: (1) First we show how to reduce POMDPs with parity objectives to POMDPs with parity objectives with priorities in $\{0, 1, 2\}$; and (2) then show how to reduce POMDPs with priorities in $\{0, 1, 2\}$ to coBüchi objectives, for almost-sure winning.

4.2.1 Almost-sure parity to almost-sure parity with three priorities

Given a POMDP $G = (S, A, \delta, \mathcal{O}, \gamma, s_0)$ and a parity objective $\mathsf{Parity}(p)$ with priority set $D = \{0, \ldots, 2 \cdot d + 1\}$, we will construct a POMDP $\overline{G} = (\overline{S}, A, \overline{\delta}, \mathcal{O}, \overline{\gamma}, \overline{s}_0)$ together with a parity objective $\mathsf{Parity}(\overline{p})$ which assigns priorities from the set $\{0, 1, 2\}$ such that almost-sure winning under finite-memory strategies is preserved. Let I denote the set $\{0, \ldots, d\}$. Intuitively to construct $\overline{G}$ we form a copy $\overline{G}_i$ of the POMDP G for each $i \in I$. The game starts in the initial state of the copy $\overline{G}_d$. The transition probabilities in the copy $\overline{G}_i$ are as follows: for states with priority at least $2 \cdot i$ it mimics the transition of G; and for states with priority strictly smaller than $2 \cdot i$ it mimics the transition of G with probability $1/2$ and with probability $1/2$ moves to the copy $i - 1$ (i.e., to $\overline{G}_{i-1}$). In $\overline{G}_i$, states with priority $2 \cdot i$ and $2 \cdot i + 1$ are assigned priorities 0 and 1, respectively, and all other states are assigned priority 2. We now present the formal construction of $\overline{G}$:

- $\overline{S} = S \times I$

- We define the transition function $\overline{\delta}$ in two steps; for a state $(s, i) \in S \times I$ and an action $a \in A$:

  1. $\overline{\delta}((s,i), a)((s', i)) = \begin{cases} \delta(s,a)(s') & \text{if } p(s) \geq 2 \cdot i \\ \frac{\delta(s,a)(s')}{2} & \text{otherwise;} \end{cases}$

  2. $\overline{\delta}((s,i), a)((s', i-1)) = \frac{\delta(s,a)(s')}{2}$ if $p(s) < 2 \cdot i$

- $\overline{\gamma}((s,i)) = \gamma(s)$;

- $\overline{s}_0 = (s_0, d)$.

The new parity objective $\mathsf{Parity}(\overline{p})$ assigning priorities $\{0, 1, 2\}$ is defined as follows:

$$\overline{p}((s,i)) = \begin{cases} 0 & \text{if } p(s) = 2 \cdot i; \\ 1 & \text{if } p(s) = 2 \cdot i + 1; \\ 2 & \text{otherwise;} \end{cases}$$

Lemma 16. *If there exists a finite-memory almost-sure winning strategy in the POMDP G for the objective* Parity(p), *then there exists a finite-memory almost-sure winning strategy with the same memory states in the POMDP $\overline{G}$ for the objective* Parity($\overline{p}$) *with three priorities.*

Proof. Let $\sigma = (\sigma_u, \sigma_n, M, m_0)$ be a finite-memory almost-sure winning strategy in the POMDP G for the objective Parity(p) and $\widehat{G}$ the Markov Chain $G \upharpoonright \sigma$. We show that the strategy σ is also almost-sure winning in the POMDP $\overline{G}$. Consider the Markov Chain $\widehat{G}' = \overline{G} \upharpoonright \sigma$. We need to show that for every recurrent class $\overline{C}$ reachable from the starting state (s_0, d, m_0) we have that $\min(\overline{p}(\text{Proj}_1(\overline{C})))$ is even to show that σ is an almost-sure winning strategy in $\overline{G}$. We will show that if there is a reachable recurrent class in $\widehat{G}'$ with minimum priority odd, then there is a reachable recurrent class in $\widehat{G}$ with minimum priority odd contradicting that σ is an almost-sure winning strategy in G.

Assume towards contradiction that there exists a recurrent class $\overline{C}$ reachable from (s_0, d, m_0) in $\widehat{G}'$ such that the minimal priority is odd, i.e., $\min(\overline{p}(\text{Proj}_1(\overline{C})))$ is odd (i.e., $\overline{C}$ contains a priority 1 state but no priority 0 state). By the construction of $\overline{G}$, for every copy $\overline{G}_i$, there are transitions only to the states in the copy $\overline{G}_i$ or to the lower copy $\overline{G}_{i-1}$. Hence there are no transitions from a lower copy to a higher copy. Hence every recurrent class in $\widehat{G}'$ must be contained in some copy. Let the recurrent class $\overline{C}$ be contained in copy i, i.e., $\overline{C} \subseteq S \times \{i\} \times M$. Also note that by construction, every state with priority strictly smaller than $2 \cdot i$ (by priority function p) has positive probability transition to a lower copy and hence such states do not belong to the recurrent class. Since $\min(\overline{p}(\text{Proj}_1(\overline{C})))$ is odd it follows that $\overline{C}$ does not contain a state with priority 0 by $\overline{p}$ (i.e., priority $2 \cdot i$ by p) but contains some state with priority 1 by $\overline{p}$ (i.e., priority $2 \cdot i + 1$ by p), i.e., (i) $\overline{C} \subseteq \left((\bigcup_{j \geq 2 \cdot i} p^{-1}(j)) \times \{i\} \times M \right)$ ($\overline{C}$ is contained in the copy $\overline{G}_i$ and the priorities of the states in $\overline{C}$ are at least $2 \cdot i$); (ii) $\overline{C} \cap (p^{-1}(2 \cdot i) \times \{i\} \times M) = \emptyset$ ($\overline{C}$ contains no priority 0 state according to $\overline{p}$); and (iii) $\overline{C} \cap (p^{-1}(2 \cdot i + 1) \times \{i\} \times M) \neq \emptyset$ ($\overline{C}$ contains some priority 1 state according to $\overline{p}$). Observe that due to the definition of observations whenever a state (s, i, m) is reachable in $\widehat{G}'$ we have that the state (s, m) is also reachable in $\widehat{G}$ (since memories of the strategies are the same and the observation function cannot distinguish between copies). It follows that the set of states $C = \{(s, m) \mid (s, i, m) \in \overline{C}\}$ are reachable from (s_0, m_0) in $\widehat{G}$. Moreover as transition probabilities for states (s, j) with $j \geq 2 \cdot i$ are not modified in the copy $\overline{G}_i$ it follows that C is a recurrent class reachable in $\widehat{G}$. Thus we have a recurrent class C reachable from (s_0, m_0) in $\widehat{G}$ such that $C \cap (p^{-1}(2 \cdot i) \times M) = \emptyset$ and $C \cap (p^{-1}(2 \cdot i + 1) \times M) \neq \emptyset$, i.e., the minimum priority is $2 \cdot i + 1$ and odd. This contradicts that σ is an almost-sure winning strategy in G for Parity(p). Hence it follows σ is an almost-sure winning strategy in $\overline{G}$ for Parity($\overline{p}$) and the desired result follows. $\square$

Lemma 17. *If there exists a finite-memory almost-sure winning strategy in the POMDP $\overline{G}$ for the objective* Parity($\overline{p}$) *with three priorities, then there exists a finite-memory almost-sure winning strategy with the same memory states in the POMDP G for the objective* Parity(p).

Proof. As in the previous lemma let $\sigma = (\sigma_u, \sigma_n, M, m_0)$ be a finite-memory almost-sure winning strategy in the POMDP $\overline{G}$ for the objective Parity($\overline{p}$) and $\widehat{G}'$ the Markov Chain $\overline{G} \upharpoonright \sigma$. We show that the strategy σ is also almost-sure winning in the POMDP G. We consider the Markov Chain $\widehat{G} = G \upharpoonright \sigma$. We need to show that for all recurrent classes C reachable in $\widehat{G}$ from (s_0, m_0) the minimal priority is even.

Assume towards contradiction that there exists a reachable recurrent class C from (s_0, m_0) in $\widehat{G}$ with minimal priority odd, and let the minimal priority be $2 \cdot i + 1$. We need to show that this implies that there exists a reachable recurrent class from (s_0, d, m_0) in $\widehat{G}'$ with minimal priority odd (as we consider only priorities $0, 1, 2$, the minimal priority is 1). Consider the subset of states $\overline{C} = \{(s, i, m) \mid (s, m) \in C\}$. The minimal priority of the set in $\overline{G}$ is 1 since C does not contain any state with priority strictly smaller $2 \cdot i + 1$ and has some state with priority $2 \cdot i + 1$. The transition function $\overline{\delta}$ matches the transition function δ on states of $\text{Proj}_1(C)$ for any action $a \in A$. Therefore, $\overline{C}$ forms a recurrent class in $\widehat{G}'$. It remains to show that $\overline{C}$

24

is reachable from the initial state of $\widehat{G}'$. Let (s, i, m) be a state in $\overline{C}$ such that $\overline{p}(s) = 1$: the state (s, m) is reachable in $\widehat{G}$ from (s_0, m_0) since (s, m) is a state in the recurrent class C reachable from (s_0, m_0) in $\widehat{G}$. Then for the starting copy $\overline{G}_d$ we have that (s, d, m) is reachable from (s_0, d, m_0) in $\widehat{G}'$. We now present two simple facts:

1. For all states $(s', m') \in C$ we have that (s, m) is reachable from (s', m') in $\widehat{G}$ (since C is a recurrent class and both (s', m') and (s, m) belong to C), and it follows that for all $j \in I$ we have that (s, j, m) is reachable from (s', j, m') in the copy j.

2. Since $\overline{p}((s, i, m)) = 1$ we have that $p(s) = 2 \cdot i + 1$, and for all $j > i$, in $\overline{G}_j$ if the state (s, j, m) is reached, then with positive probability we reach the copy $j - 1$ (some state $(s', j - 1, m')$). Moreover, since $(s, m) \in C$, for all $j > i$, from (s, j, m) we reach a state $(s', j - 1, m')$ such that $(s', m') \in C$.

From the above two facts it follows that for all $j > i$ we have $(s, j - 1, m)$ is reachable from (s, j, m). It follows that (s, i, m) is reachable from (s, d, m) and since (s, d, m) is reachable from (s_0, d, m_0) it follows that (s, i, m) is reachable from (s_0, d, m_0). Hence $\overline{C}$ is reachable from (s_0, d, m_0) and we have a contradiction to the fact that σ is an almost-sure winning strategy in $\overline{G}$ for $\mathsf{Parity}(\overline{p})$. It follows that σ is an almost-sure winning strategy in G for $\mathsf{Parity}(p)$ and the desired result follows. $\qquad\square$

Lemma 16 and Lemma 17 gives us the following lemma.

Lemma 18. *If there exists a finite-memory almost-sure winning strategy σ in the POMDP G with the objective* $\mathsf{Parity}(p)$, *then σ is an almost-sure winning strategy in the POMDP $\overline{G}$ with the objective* $\mathsf{Parity}(\overline{p})$ *with three priorities; and vice versa.*

Next we show how to reduce the problem of almost-sure winning for parity objectives with priorities from the set $\{0, 1, 2\}$ to the problem of almost-sure winning for coBüchi objectives in POMDPs.

4.2.2 Almost-sure parity with three priorities to almost-sure coBüchi

Consider a POMDP $\overline{G} = (\overline{S}, A, \overline{\delta}, \mathcal{O}, \overline{\gamma}, \overline{s}_0)$ with a parity objective $\mathsf{Parity}(\overline{p})$ assigning priorities from the set $\{0, 1, 2\}$. We construct a POMDP $\widetilde{G} = (\widetilde{S}, A, \widetilde{\delta}, \mathcal{O}, \widetilde{\gamma}, \widetilde{s}_0)$ with a coBüchi objective $\mathsf{coBuchi}(\widetilde{T})$, where the set of states $\widetilde{T}$ is going to be defined as $\widetilde{p}^{-1}(2)$ for a function $\widetilde{p}$ assigning priorities from the set $\{1, 2\}$. Intuitively, for states with priority 1 and 2, the transition function $\widetilde{\delta}$ mimics the transitions of $\overline{\delta}$; and for states with priority 0, the transition function $\widetilde{\delta}$ mimics the transitions of $\overline{\delta}$ with probability $1/2$ and with probability $1/2$ goes to a newly added absorbing state that is assigned priority 2. Formally the POMDP $\widetilde{G}$ is defined as follows:

- $\widetilde{S} = \overline{S} \cup \{\widetilde{s}_r\}$;

- $\widetilde{\delta}$ is defined for all states $s \in \overline{S}$ and all actions $a \in A$ as follows:

  1. $\widetilde{\delta}(s, a)(s') = \begin{cases} \overline{\delta}(s, a)(s') & \text{if } \overline{p}(s) \in \{1, 2\}; \\ \frac{\overline{\delta}(s,a)(s')}{2} & \text{if } \overline{p}(s) = 0; \end{cases}$

  2. $\overline{\delta}(s, a)(\widetilde{s}_r) = 1/2$ if $\overline{p}(s) = 0$;

  3. $\overline{\delta}(\widetilde{s}_r, a)(\widetilde{s}_r) = 1$, i.e., $\widetilde{s}_r$ is an absorbing state;

- $\widetilde{\mathcal{O}} = \mathcal{O} \cup \{\widetilde{s}_r\}$, i.e., the additional state is a new observation;

- $\widetilde{\gamma}(s) = \begin{cases} \overline{\gamma}(s) & \text{if } s \in \overline{S} \\ \widetilde{s}_r & \text{if } s = \widetilde{s}_r \end{cases}$

The coBüchi objective is defined by a priority function $\widetilde{p}$, that is defined as:

$$\widetilde{p}(s) = \begin{cases} \overline{p}(s) & \text{if } \overline{p}(s) \in \{1, 2\} \\ 2 & \text{if } \overline{p}(s) = 0 \text{ or } s = \widetilde{s}_r \end{cases}$$

The objective in $\widetilde{G}$ is $\mathsf{coBuchi}(\widetilde{p}^{-1}(2))$.

Lemma 19. *If there exists a finite-memory almost-sure winning strategy σ in the POMDP $\overline{G}$ with the objective* $\mathsf{Parity}(\overline{p})$ *with three priorities, then σ is an almost-sure winning strategy in the POMDP $\widetilde{G}$ with the objective* $\mathsf{coBuchi}(\widetilde{p}^{-1}(2))$ *and vice versa.*

Proof. We start with the first direction. Let $\sigma = (\sigma_u, \sigma_n, M, m_0)$ be a finite-memory almost-sure winning strategy in $\overline{G}$, we claim that σ is also almost-sure winning in $\widetilde{G}$. Assume towards contradiction that there exists a reachable recurrent class C in the Markov chain $\widetilde{G} \upharpoonright \sigma$ such that the minimal priority in the class is 1. Then C cannot contain the newly added absorbing state $\widetilde{s}_r$, as $\widetilde{p}(\widetilde{s}_r) = 2$ and if a recurrent class contains the absorbing state $\widetilde{s}_r$, then it contains only the state $\widetilde{s}_r$ as there is no edge going from $\widetilde{s}_r$ to a different state in the POMDP $\widetilde{G}$. It follows that the set C is reachable in $\overline{G} \upharpoonright \sigma$, and due to the definition of the transition functions forms a recurrent class. Since C contains a state s with priority $\widetilde{p}(s) = 1$, we have that $\overline{p}(s)$ is also 1, so there is a state with priority 1 in C. It remains to rule out the possibility that C contains states s' with priority $\overline{p}(s') = 0$, but that follows from the fact that whenever there was a state with priority 0, no matter what action was played, there was a positive probability of reaching $\widetilde{s}_r$. So C contains a state with priority 1 and all the other states have priority 1 or 2. It follows that there exists a reachable recurrent class in $\overline{G} \upharpoonright \sigma$, where the minimal priority is odd. This contradicts our assumption that σ is almost-sure winning in $\overline{G}$.

In the second direction assume that no finite-memory strategy is almost-sure winning in $\overline{G}$. Therefore, for every finite-memory strategy σ there exists a reachable recurrent class C in the Markov Chain $\overline{G} \upharpoonright \sigma$, such that the minimal priority in the class is 1, i.e., there exists a state with priority 1 and there are no states with priority 0 in C. In the Markov Chain $\widetilde{G} \upharpoonright \sigma$ the transition functions $\widetilde{\delta}$ allows every transition available in $\overline{\delta}$. It follows that the set C is reachable with positive probability in $\widetilde{G} \upharpoonright \sigma$. Since there is no state with priority 0 in C it follows that the transition function $\widetilde{\delta}$ matches the transition function $\overline{\delta}$ for all states in C and all actions $a \in A$. It follows that C is a recurrent class in the POMDP $\widetilde{G} \upharpoonright \sigma$. As all the priorities of the states in C are preserved in the priority function $\widetilde{p}$, there exists a reachable recurrent class with minimal priority 1. It follows that there is no finite-memory almost-sure winning strategy in the POMDP $\widetilde{G}$. $\qquad\square$

Lemma 18 and Lemma 19 establish item (2) of Theorem 3.

5 Computational Complexity for Parity Objectives

In this section we will present an exponential time algorithm to solve almost-sure winning in POMDPs with coBüchi objectives under finite-memory strategies (and the polynomial time reduction of Section 4 for parity objectives to coBüchi objectives allows our results to carry over to parity objectives). The results for positive Büchi is similar and the almost similar proof is omitted. The naive algorithm would be to enumerate over all finite-memory strategies with memory bounded by $2^{6 \cdot |S|}$, this leads to an algorithm that runs in double-exponential time. Instead our algorithm consists of two steps: (1) first constructing a special

kind of POMDPs where the restriction is to randomized memoryless strategies; and (2) then show how to solve the special kind of POMDPs in time polynomial in the size of the POMDP. We first introduce the special kind of POMDPs which we call belief-observation POMDPs which intuitively satisfy that the current belief is always the set of states with current observation.

Definition 7 (Belief-observation POMDP). *A POMDP $G = (S, A, \delta, \mathcal{O}, \gamma, s_0)$ is a* belief-observation POMDP *iff for every finite prefix $w = (s_0, a_0, s_1, a_1, \ldots, s_n)$ with the observation sequence $\rho = \gamma(w)$, the belief $\mathcal{B}(\rho)$ is equal to the set of states with the observation $\gamma(s_n)$, i.e., $\mathcal{B}(\rho) = \{s \in S \mid \gamma(s) = \gamma(s_n)\}$. In other words, belief-observation POMDPs are the special class of POMDPs where the current belief can be directly obtained from the current observation.*

5.1 Construction of belief-observation POMDPs for finite-memory strategies

POMDPs to belief-observation POMDPs. The goal of this section is given a POMDP G with a coBüchi objective $\mathsf{coBuchi}(p^{-1}(2))$, and a priority function with priority set $\{1, 2\}$, to construct a belief-observation POMDP $\widehat{G}$ such that if there exists a finite-memory almost-sure winning strategy in G, then there exists a randomized memoryless almost-sure winning strategy in $\widehat{G}$ for another coBüchi objective $\mathsf{coBuchi}(\widehat{p}^{-1}(2))$ and vice-versa. Since we are interested in coBüchi objectives, for the sequel of this section we will denote by $M = 2^S \times \{0, 1\}^{|S|} \times \mathfrak{D}^{|S|}$, i.e., all the possible beliefs $\mathcal{B}$, BoolRec and SetRec functions (recall that $\mathfrak{D}$ is $\mathcal{P}(\mathcal{P}(\{1, 2\}))$ for coBüchi objectives). If there exists a finite-memory almost-sure winning strategy σ, then the projected strategy $\sigma' = proj(\sigma)$ is also a finite-memory almost-sure winning strategy (by Theorem 2) and will use memory $M' \subseteq M$. The size of the constructed POMDP $\widehat{G}$ will be exponential in the size of the original POMDP G and polynomial in the size of the memory set M (and $|M| = 2^{6 \cdot |S|}$ is exponential in the size of the POMDP G). We define the set $M_{\mathsf{coBuchi}} \subseteq M$ as the memory elements, where for all states s in the belief component of the memory, the set $\mathsf{SetRec}(s)$ contains only a set with priority two, i.e., there is no state with priority 1 in the reachable recurrent classes according to SetRec. Formally,

$$M_{\mathsf{coBuchi}} = \{(Y, B, L) \in M \mid \text{ for all } s \in Y \text{ we have } L(s) = \{\{2\}\}\}$$

Construction of the new POMDP. Given a POMDP $G = (S, A, \delta, \mathcal{O}, \gamma, s_0)$ with a coBüchi objective $\mathsf{coBuchi}(p^{-1}(2))$, represented by priority function $p : S \rightarrow \{1, 2\}$, we construct a new POMDP $\widehat{G} = (\widehat{S}, \widehat{A}, \widehat{\delta}, \widehat{\mathcal{O}}, \widehat{\gamma}, \widehat{s}_0)$ with a coBüchi objective $\mathsf{coBuchi}(\widehat{p}^{-1}(2))$, for some priority function $\widehat{p}$ assigning to states in $\widehat{S}$ priorities from the set $\{1, 2\}$. We refer to the newly constructed POMDP $\widehat{G}$ as $\mathsf{AlmostCoBuchiRed}(G)$.

- The set of states $\widehat{S} = \widehat{S}_a \cup \widehat{S}_m \cup \{\widehat{s}_0, \widehat{s}_b\}$, will consist of *action-selection* states $\widehat{S}_a \subseteq S \times M$; *memory-selection* states $\widehat{S}_m \subseteq S \times 2^S \times M \times A$; $\widehat{s}_0$ is an additional initial state; and the state $\widehat{s}_b$ is a new absorbing state.

- The observation set is as follows: $\widehat{\mathcal{O}} = (M) \cup (2^S \times M \times A) \cup \{\widehat{s}_0\} \cup \{\widehat{s}_b\}$.

- The initial state of the POMDP is $\widehat{s}_0$.

- The observation mapping is defined naturally $\widehat{\gamma}((s, m)) = m$, $\widehat{\gamma}((s, Y, m, a)) = (Y, m, a)$, $\widehat{\gamma}(\widehat{s}_0) = \{\widehat{s}_0\}$, and $\widehat{\gamma}(\widehat{s}_b) = \{\widehat{s}_b\}$. In other words, except the states $\widehat{s}_0$ and $\widehat{s}_b$ the strategy cannot observe the first component of the state.

- The actions are $\widehat{A} = A \cup M$, i.e., the actions from the POMDP G or memory elements from the set M.

- We define the transition function $\widehat{\delta}$ in the following steps. First we will introduce a notion of *allowed actions*, observe that for the computation of almost-sure winning under finite-memory strategies the precise transition probabilities do not matter and therefore in the following step we will specify only the edges of the POMDP graph, and all transition probabilities are uniform over the support set.

  We call an action $a \in A$ *allowed* in observation $(Y, B, L) \in \widehat{\mathcal{O}}$ if for all states $\widehat{s} \in Y$, there exists a set $Z_\infty \subseteq \{1, 2\}$ such that if $B(\widehat{s}) = 1$, $L(\widehat{s}) = \{Z_\infty\}$, and $p(\widehat{s}) \in Z_\infty$, then for all states $\widehat{s}' \in \mathrm{Supp}(\delta(\widehat{s}, a))$ we have $p(\widehat{s}') \in Z_\infty$. Intuitively this condition enforces that once a state that corresponds to a pseudo-recurrent state is reached in the POMDP $\widehat{G}$ in the next step only states with priority in the set Z_∞ can be visited. Similarly we call an action $(Y', B', L') \in M$ *allowed* in observation $(Y', (Y, B, L), a)$ if both of the following conditions are satisfied: (i) for all states $\widehat{s} \in Y$, if $B(\widehat{s}) = 1$, then for all states $\widehat{s}' \in \mathrm{Supp}(\delta(\widehat{s}, a))$ we have that $B'(\widehat{s}') = 1$, intuitively the condition says that if the $\mathsf{BoolRec}$ function is set to 1, then for all successors the $\mathsf{BoolRec}$ function remains 1 (recall by Lemma 7 fifth point the property is ensured for projected strategies); and (ii) if $\widehat{s} \in S$ and $\widehat{s}' \in \mathrm{Supp}(\delta(\widehat{s}, a))$, we have that $L'(\widehat{s}') \subseteq L(\widehat{s})$. Intuitively the condition says the function SetRec must not increase with respect to set inclusion along the successors (recall by Lemma 7 third point the property is ensured for projected strategies).

  1. $\widehat{s}_0 \xrightarrow{m} (s_0, m)$ for all $m \in M_{\mathsf{coBuchi}} \cap \{(\{s_0\}, B, L) \mid B \in \{0, 1\}^S, L \in \mathfrak{D}^S\}$, i.e., from the initial state all memory elements from M_{coBuchi} that are consistent with the starting state can be chosen (in other words, it consists of all the ways a projected strategy of a finite-memory almost-sure winning strategy could start);

  2. $(s, (Y, B, L)) \xrightarrow{a} (s', Y', (Y, B, L), a)$ iff all of the following conditions are satisfied:
     - $s' \in \mathrm{Supp}(\delta(s, a))$; and
     - Y' is the belief update in POMDP G from belief Y under observation $o = \gamma(s')$ and action a, i.e., $Y' = \bigcup_{\widehat{s} \in Y} \mathrm{Supp}(\delta(\widehat{s}, a)) \cap \gamma^{-1}(o)$; and
     - action a is allowed in observation (Y, B, L).

  3. If an action a is not allowed in observation (Y, B, L), then we add a transition $(s, (Y, B, L)) \xrightarrow{a} \widehat{s}_b$, i.e., if the conditions are not satisfied the action leads to the state $\widehat{s}_b$ that will be a loosing absorbing state in the POMDP $\widehat{G}$.

  4. $(s', Y', (Y, B, L), a) \xrightarrow{(Y', B', L')} (s', (Y', B', L'))$ iff the action (Y', B', L') is allowed in the observation $(Y', (Y, B, L), a)$. Again if an action is not allowed, then the transition leads only to $\widehat{s}_b$.

  5. The state $\widehat{s}_b$ is an absorbing state, i.e., $\widehat{s}_b \xrightarrow{\widehat{a}} \widehat{s}_b$ for all actions $\widehat{a} \in \widehat{A}$.

Intuitively, $\widehat{G}$ allows all possible ways that a projected strategy of a finite-memory almost-sure winning strategy could possibly play in G. We define the coBüchi objective $\mathsf{coBuchi}(\widehat{p}^{-1}(2))$ with the priority function for the POMDP $\widehat{G}$ as $\widehat{p}((s, m)) = \widehat{p}((s, Y, m, a)) = p(s)$. The priority for the initial state $\widehat{p}(\widehat{s}_0)$ may be set to an arbitrary priority from $\{1, 2\}$ as the initial state will be visited only once. The priority for the state $\widehat{s}_b$ is set to 1, i.e., $\widehat{p}(\widehat{s}_b) = 1$. We will refer to the above construction as $\mathsf{AlmostCoBuchiRed}$ construction, i.e., $\widehat{G} = \mathsf{AlmostCoBuchiRed}(G)$. We first argue that $\widehat{G}$ is a belief-observation POMDP.

Lemma 20. *The POMDP $\widehat{G}$ is a belief-observation POMDP.*

Proof. Note that the observations are defined in a way that the first component cannot be observed. Given a sequence $\widehat{w}$ of states and actions in $\widehat{G}$ with the observation sequence $\widehat{\rho} = \widehat{\gamma}(\widehat{w})$ we will show that the possible first components of the states in the belief $\mathcal{B}(\widehat{\rho})$ are equal the updated belief components Y' in the observation. Intuitively the proof holds as the Y component is the belief and the belief already represents exactly the set of states in which the POMDP can be with positive probability. We now present the formal argument. Let us denote by $\mathsf{Proj}_1(\mathcal{B}(\widehat{\rho})) \subseteq S$ the projection on the first component of the states in the belief. One inclusion is trivial since for every reachable state $(s, (Y, B, L))$ we have $s \in Y$ (resp. for states $(s', Y', (Y, B, L), a)$ we have that $s' \in Y'$). Therefore we have $\mathsf{Proj}_1(\mathcal{B}(\widehat{\rho})) \subseteq Y$ (resp. Y').

We prove the second inclusion by induction with respect to the length of the play prefix:

- **Base case:** We show the base case for prefixes of length 1 and 2. The first observation is always $\{\widehat{s}_0\}$ which contains only a single state, so there is nothing to prove. Similarly the second observation in the game is of the form $(\{s_0\}, B, L)$ for some $B \in \{0, 1\}^S, L \in \mathfrak{D}^S$, and the argument is the same.

- **Induction step:** Let us a consider a prefix $\widehat{w}' = \widehat{w} \cdot a \cdot (s', Y', (Y, B, L), a)$ where $a \in A$ and the last transition is $(s, (Y, B, L)) \xrightarrow{a} (s', Y', (Y, B, L), a)$ in the POMDP $\widehat{G}$. By induction hypothesis we have that $\mathcal{B}(\widehat{\gamma}(\widehat{w})) = \{(s, (Y, B, L)) \mid s \in Y\}$. The new belief is computed (by definition) as

$$\mathcal{B}(\widehat{\gamma}(\widehat{w}')) = \bigcup_{\widehat{s} \in Y} \mathsf{Supp}(\widehat{\delta}((\widehat{s}, (Y, B, L)), a)) \cap \widehat{\gamma}^{-1}((Y', (Y, B, L), a)).$$

  Let $s_{Y'}$ be a state in Y', we want to show that $(s_{Y'}, Y', (Y, B, L), a)$ is in $\mathcal{B}(\widehat{\gamma}(\widehat{w}'))$. Due to the definition of the belief update there exists a state s_Y in Y such that $s_Y \xrightarrow{a} s_{Y'}$ and $(s_Y, (Y, B, L)) \in \mathcal{B}(\widehat{\gamma}(\widehat{w}))$. As $\gamma(s_{Y'}) = \gamma(s')$, it follows that $(s_{Y'}, Y', (Y, B, L), a) \in \bigcup_{\widehat{s} \in \mathcal{B}(\widehat{\gamma}(\widehat{w}))} \mathsf{Supp}(\widehat{\delta}(\widehat{s}, a))$ and as $(s_{Y'}, Y', (Y, B, L), a) \in \widehat{\gamma}^{-1}((Y', (Y, B, L), a))$, the result follows.

  The case when the prefix is extended with an memory action $m \in M$ is simpler as the first two components do not change during the transition.

The desired result follows. $\qquad\square$

The proof of the following two lemmas will use some desired properties of the projected strategy of a finite-memory strategy and the BoolRec and SetRec functions established in Section 3. The properties are as follows:

1. *(Property A for BoolRec functions).* For every run of the Markov chain obtained from the POMDP and a projected strategy $proj(\sigma)$ of a finite-memory strategy σ, if BoolRec is set to 1, then for all successors BoolRec remains 1 (follows from the fifth point of Lemma 7).

2. *(Property B for SetRec functions).* For every run of the Markov chain obtained from the POMDP and a projected strategy $proj(\sigma)$ of a finite-memory strategy σ, SetRec functions are non-increasing along the steps of the run (follows from the third point of Lemma 7).

3. *(Property C for BoolRec and SetRec functions).* For every run of the Markov chain obtained from the POMDP and a projected strategy $proj(\sigma)$ of a finite-memory strategy σ, if BoolRec is set to 1 for a state s, then all reachable states from that point have a priority in $\mathsf{SetRec}_{proj(\sigma)}(s)$ (follows from the sixth point of Lemma 7).

Lemma 21. *If there exists a finite-memory almost-sure winning strategy in the POMDP G for the coBüchi objective* coBuchi($p^{-1}(2)$), *then there exists a randomized memoryless almost-sure winning strategy in the belief-observation POMDP $\widehat{G}$ for the coBüchi objective* coBuchi($\widehat{p}^{-1}(2)$).

Proof. Assume there exists a finite-memory almost-sure winning strategy σ, then by Theorem 2 there exists a finite-memory almost-sure winning strategy $\sigma' = proj(\sigma)$, which uses memory $M' \subseteq M$. Let $\sigma' = (\sigma'_u, \sigma'_n, M', (\{s_0, \}, B_0, L_0))$ be the almost-sure winning strategy in the POMDP G. We fix the strategy σ' in the POMDP G and obtain a Markov Chain $G_1 = G \upharpoonright \sigma'$. We define a randomized memoryless observation-based strategy $\widehat{\sigma} : \widehat{\mathcal{O}} \to \mathcal{D}(\widehat{A})$ in the POMDP $\widehat{G}$ as follows:

- The deterministic action in the initial observation is: $\widehat{\sigma}(\{\widehat{s}_0\}) = (\{s_0\}, B_0, L_0)$.

- In the action-selection observation (Y, B, L) we define $\widehat{\sigma}((Y, B, L)) = \sigma'_n((Y, B, L))$.

- In the memory-selection observation $(Y', (Y, B, L), a)$ we define $\widehat{\sigma}((Y', (Y, B, L), a))$ to play uniformly actions from the set $\mathrm{Supp}((\sigma'_u((Y, B, L), o', a)))$, where o' is the unique observation such that all states in Y' have observation o' in G.

- In the observation of the absorbing state $\{\widehat{s}_b\}$ no matter what actions are played the state $\widehat{s}_b$ is not left.

We fix the memoryless strategy $\widehat{\sigma}$ in the POMDP $\widehat{G}$ and obtain a Markov chain $G_2 = \widehat{G} \upharpoonright \widehat{\sigma}$. For simplicity we will write $(s, (Y, B, L)) \to (s', (Y', B', L'))$ whenever $(s, (Y, B, L)) \to (s', Y', (Y, B, L), a) \to (s', (Y', B', L'))$ for some $a \in A$ in G_2. Note that omitting the intermediate state does not affect the objective as $\widehat{p}((s', Y', (Y, B, L), a)) = \widehat{p}((s', (Y', B', L')))$.

The strategy $\widehat{\sigma}$ will in the first step select the $(\{s_0\}, B_0, L_0)$ action and reach state $(s_0, (\{s_0\}, B_0, L_0))$ in G_2. We will show that the two Markov chains reachable from initial state $(s_0, (\{s_0\}, B_0, L_0))$ in G_1 and $(s_0, (\{s_0\}, B_0, L_0))$ in G_2 are isomorphic (when considering the simplified edges in G_2).

- Let $(s, (Y, B, L)) \to (s', (Y', B', L'))$ be an edge in G_1, then there exists (i) an action $a \in \mathrm{Supp}(\sigma'_n((Y, B, L)))$, such that $s' \in \mathrm{Supp}(\delta(s, a))$, (ii) Y' is the belief update from Y under observation $\gamma(s')$ and action a, and (iii) $(Y', B', L') \in \mathrm{Supp}(\sigma'_u((Y, B, L), \gamma(s'), a))$. First we show that there is a transition $(s, (Y, B, L)) \xrightarrow{a} (s', Y', (Y, B, L), a)$ in the POMDP $\widehat{G}$. We verify three properties of the transition functions $\widehat{\delta}$: The property that $s' \in \mathrm{Supp}(\delta(s, a))$ and that Y' is the belief update follows from the facts (i) and (ii) mentioned above. Next we show that action a is allowed in observation (Y, B, L), i.e., we need to verify that for all states $\widehat{s} \in Y$ such that $B(\widehat{s}) = 1$, and there is a subset of priorities $Z_\infty \subseteq \{1, 2\}$ such that $L(\widehat{s}) = \{Z_\infty\}$, and $p(\widehat{s}) \in Z_\infty$ we have that all states reachable in one step $\widehat{s}' \in \mathrm{Supp}(\delta(\widehat{s}, a))$ satisfy $p(\widehat{s}') \in Z_\infty$. Consider an arbitrary state $\widehat{s} \in Y$ such that $B(\widehat{s}) = 1$, $L(\widehat{s}) = \{Z_\infty\}$ and $p(\widehat{s}) \in Z_\infty$. Note that $(\widehat{s}, (Y, B, L))$ is a reachable pseudo-recurrent state in the Markov chain G_1 since $B(\widehat{s}) = 1$, $L(\widehat{s}) = \{Z_\infty\}$ and $p(\widehat{s}) \in Z_\infty$. As the strategy σ' is almost-sure winning it follows that all recurrent classes reachable from $(\widehat{s}, (Y, B, L))$ contain only states with priority 2 i.e., $Z_\infty = \{2\}$. Moreover by property A and C it follows that only states with priority in Z_∞ (i.e, with priority 2) are reachable from $(\widehat{s}, (Y, B, L))$ in G_1. It follows that action a is allowed in state $(s, (Y, B, L))$. By the definition of the strategy $\widehat{\sigma}$ this action is played with positive probability and therefore $(s, (Y, B, L)) \to (s', Y', (Y, B, L), a)$ is an edge in G_2. Similarly, we show that there is a transition $(s', Y', (Y, B, L), a) \xrightarrow{(Y', B', L')} (s', (Y', B', L'))$ in $\widehat{G}$. To show that (Y', B', L') is an allowed action in observation $(Y', (Y, B, L), a)$ we consider all states $\widehat{s} \in Y$, such that $B(\widehat{s}) = 1$, and all reachable states $\widehat{s}' \in \mathrm{Supp}(\delta(\widehat{s}, a))$, and want to

show that $B'(\widehat{s}') = 1$. As the state $(\widehat{s}, (Y, B, L))$ is reachable and σ' is a projected strategy of an almost-sure winning strategy, it follows by property B and C of BoolRec and SetRec functions of the projected strategy σ', i.e., point three and five in Lemma 7, that all the memories in the memory update $\mathrm{Supp}(\sigma'_u((Y, B, L), \gamma(s'), a))$ of the projected strategy satisfy that $B'(\widehat{s}') = 1$. The second property of the non-increasing SetRec function is proved similarly. It follows that the action satisfies the requirements of item (4) of the transition function $\widehat{\delta}$. As before (Y', B', L') is played with positive probability and hence $(s', Y', (Y, B, L), a) \rightarrow (s', (Y', B', L'))$ is an edge in G_2. It follows that $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ is an edge in the simplified graph of G_2.

- In the other direction let us consider an edge $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ in the simplified graph of G_2, it follows that there exists an action a, such that there are edges $(s, (Y, B, L)) \rightarrow (s', Y', (Y, B, L), a) \rightarrow (s', (Y', B', L'))$ in the full graph G_2. By the definition of the POMDP $\widehat{G}$ we get that $s' \in \mathrm{Supp}(\delta(s, a))$ and Y' is the belief update from Y under observation $\gamma(s')$ and action a. As action a was played with positive probability it follows by the definition of the strategy $\widehat{\sigma}$ that $a \in \mathrm{Supp}(\sigma'_n((Y, B, L)))$ and similarly (Y', B', L') being played by $\widehat{\sigma}$ we get that $(Y', B', L') \in \mathrm{Supp}(\sigma'_u((Y, B, L), \gamma(s'), a))$. Hence we get that $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ is an edge in G_1.

The desired result follows. $\qquad\qquad\square$

Lemma 22. *If there exists a randomized memoryless almost-sure winning strategy in the belief-observation POMDP $\widehat{G}$ for the coBüchi objective $\mathrm{coBuchi}(\widehat{p}^{-1}(2))$, then there exists a finite-memory almost-sure winning strategy in the POMDP G for the coBüchi objective $\mathrm{coBuchi}(p^{-1}(2))$.*

Proof. Given a memoryless strategy $\widehat{\sigma}$ in $\widehat{G}$, we define the finite-memory strategy $\sigma = (\sigma_u, \sigma_n, M, m_0)$ in G as follows:

- $\sigma_n((Y, B, L)) = \widehat{\sigma}((Y, B, L))$;

- $\sigma_u((Y, B, L), o, a)$ update uniformly to elements from the set $\mathrm{Supp}(\widehat{\sigma}((Y', (Y, B, L), a)))$, where Y' is the belief update from Y under observation o and action a.

- $m_0 = \widehat{\sigma}(\{\widehat{s}_0\})$. Note that this can be in general a probability distribution. Since we require the initial memory to be deterministic, we can model this property by adding an additional initial state and memory state from which the required randomized memory update is performed.

We fix the finite-memory strategy σ in the POMDP G to obtain a Markov Chain $G_1 = G \upharpoonright \sigma$ and similarly fixing the memoryless strategy $\widehat{\sigma}$ in the POMDP $\widehat{G}$ to obtain a Markov Chain $G_2 = \widehat{G} \upharpoonright \widehat{\sigma}$.

As in the previous lemma we will consider a simplified graph G_2 and write $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ whenever $(s, (Y, B, L)) \rightarrow (s', Y', (Y, B, L), a) \rightarrow (s', (Y', B', L'))$ for some $a \in A$ in G_2. We show that the two graphs reachable from states $(s_0, (\{s_0\}, B_0, L_0))$ in G_1 and $(s_0, (\{s_0\}, B_0, L_0))$ in G_2 are isomorphic. Note that the absorbing state $\widehat{s}_b$ is not reachable in the Markov chain G_2, otherwise there would be reachable recurrent class in G_2 containing the state $\widehat{s}_b$ and no other state (follows from the fact that $\widehat{s}_b$ is an absorbing state). As the priority $\widehat{p}(\widehat{s}_b)$ is 1 it follows that there would be a reachable recurrent class with minimal priority 1 and contradicting the assumption that $\widehat{\sigma}$ is an almost-sure winning strategy.

- Let $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ be an edge in G_1, then there exists an action $a \in A$ such that (i) $a \in \mathrm{Supp}(\widehat{\sigma}((Y, B, L)))$, (ii) $s' \in \mathrm{Supp}(\delta(s, a))$, and (iii) $(Y', B', L') \in$

$\text{Supp}(\widehat{\sigma}(Y', (Y, B, L), a))$. Therefore there are edges $(s, (Y, B, L)) \rightarrow (s', Y', (Y, B, L), a)$ and $(s', Y', (Y, B, L), a) \rightarrow (s', (Y', B', L'))$ in G_2.

- In the other direction let there be an edge $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ in the simplified graph of G_2, then there exists an action $a \in A$ such that $(s, (Y, B, L)) \rightarrow (s', Y', (Y, B, L), a) \rightarrow (s', (Y', B', L'))$ are transitions in the full graph of G_2. By the definition of the POMDP $\widehat{G}$ and the strategy σ we get that (i) $s' \in \text{Supp}(\delta(s, a))$, (ii) $a \in \text{Supp}(\sigma_n((Y, B, L)))$ and (iii) $(Y', B', L') \in \text{Supp}(\sigma_u((Y, B, L), \gamma(s'), a))$. Therefore $(s, (Y, B, L)) \rightarrow (s', (Y', B', L'))$ is an edge in the graph of G_1.

The desired result follows. $\qquad\square$

5.2 Polytime algorithm for belief-observation POMDPs

In this section we will present a polynomial time algorithm for the computation of the almost-sure winning set for the belief-observation POMDP $\widehat{G}$ for coBüchi objectives under randomized memoryless strategies. The algorithm will use solutions of almost-sure winning sets for safety and reachability objectives.

POMDPs with available actions. For simplicity in presentation we will consider POMDPs with an available action function that maps to every observation the set of available actions for the observation, i.e., we consider POMDPs as tuples $(S, A, \delta, \mathcal{O}, \Gamma, \gamma, s_0)$, where the function $\Gamma : \mathcal{O} \rightarrow 2^A \setminus \emptyset$ maps every observation to a non-empty set of available actions. Note that this is for simplicity in presentation, as if an action is not available for an observation, then a new state can be added that is loosing and for every unavailable action transitions can be added to the newly added loosing state (thus making all actions available).

Almost-sure winning observations. For an objective φ, we denote by $\text{Almost}(\varphi) = \{o \in \mathcal{O} \mid$ there exists a randomized memoryless strategy σ such that for all $s \in \gamma^{-1}(o)$. $\mathbb{P}_s^\sigma(\varphi) = 1\}$ the set of observations such that there is a randomized memoryless strategy to ensure winning with probability 1 from all states of the observation. Our goal is to compute $\text{Almost}(\text{coBuchi}(\widehat{p}^{-1}(2)))$. Also note that since we consider belief-observation POMDPs we can only consider beliefs that correspond to all states of an observation. First we introduce one necessary notation:

- *(Allow).* Given a set $O \subseteq \mathcal{O}$ of observations and an observation $o \in O$ we define by $\text{Allow}(o, O)$ the set of actions that when played in o ensures that the next observation is in O, i.e., more formally:

$$\text{Allow}(o, O) = \{a \in \Gamma(o) \mid \bigcup_{s \in \gamma^{-1}(o)} \gamma(\text{Supp}(\delta(s, a))) \subseteq O\}.$$

We will consider the POMDP $\widehat{G} = \text{AlmostCoBuchiRed}(G)$ obtained by the construction for reduction to belief-observation POMDPs.

Definition 8. *Given the POMDP $\widehat{G}$, for a set $F \subseteq \widehat{S}$ of states, if $\{s_0\} \in \text{Almost}(\text{Safe}(F))$, we define a POMDP $\widehat{G}_{\text{Safe}(F)} = (\widehat{S}, A, \delta, \text{Almost}(\text{Safe}(F)), \widehat{\Gamma}, \widehat{\gamma}, s_0)$ as follows:*

- *The set of states is $\widehat{S} = \gamma^{-1}(\text{Almost}(\text{Safe}(F)))$;*

- *the available actions are restricted as follows: $\widehat{\Gamma}(o) = \text{Allow}(o, \text{Almost}(\text{Safe}(F)))$; and*

- *the observation mapping function $\widehat{\gamma}(s) = \gamma(s)$.*

Lemma 23. *The POMDP $\widehat{G}_{\mathsf{Safe}(F)}$ is a belief-observation POMDP.*

Proof. Follows directly from the fact that $\widehat{G}$ is a belief-observation POMDP. $\square$

Almost-sure winning for coBüchi objectives. In this part we will show how to decide whether an observation $o \in \widehat{\mathcal{O}}$ is an almost-sure winning observation for the coBüchi objective $\mathsf{coBuchi}(\widehat{p}^{-1}(2))$ in the belief-observation POMDP $\widehat{G}$ (i.e., whether $o \in \mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2)))$). We will show that the computation can be achieved by computing almost-sure winning regions for safety and reachability objectives. The steps of the computation are as follows:

1. *(Step 1).* Let $F = \widehat{S} \setminus \widehat{s}_b$ and we first compute $\widehat{G}_{\mathsf{Safe}(F)}$. This step requires the computation of the almost-sure winning for safety objectives.

2. *(Step 2).* Let $\widehat{S}_{wpr} \subseteq \widehat{S}$ denote the subset of states that intuitively correspond to *winning pseudo-recurrent (wpr)* states, i.e., formally it is defined as follows:

$$\widehat{S}_{wpr} = \{(s, (Y, B, L)) \mid B(s) = 1, L(s) = \{\{2\}\} \text{ and } \widehat{p}(s) = 2\}.$$

In the restricted POMDP $\widehat{G}_{\mathsf{Safe}(F)}$ we compute the set of observations $W_2 = \mathsf{Almost}(\mathsf{Reach}(\widehat{S}_{wpr}))$. We will show that $W_2 = \mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2)))$. This step requires the computation of the almost-sure winning for reachability objectives.

In the following two lemmas we establish the two required inclusions to show $W_2 = \mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2)))$.

Lemma 24. $W_2 \subseteq \mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2)))$.

Proof. Let $o \in W_2$ be an observation in W_2, and we show how to construct a randomized memoryless almost-sure winning strategy ensuring that $o \in \mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2)))$. Let σ be the strategy produced by the computation of $\mathsf{Almost}(\mathsf{Reach}(\widehat{S}_{wpr}))$. We will show that the same strategy ensures also $\mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2)))$. As in every observation o the strategy σ plays only a subset of actions that are in $\mathsf{Allow}(o, \mathsf{Almost}(\mathsf{Safe}(F)))$ (to ensure safety in F), where $F = \widehat{S} \setminus \widehat{s}_b$, the absorbing state $\widehat{s}_b$ is not reachable. Also with probability 1 the set $\widehat{S}_{wpr}$ is reached. We show that for all states $(s, (Y, B, L)) \in \widehat{S}_{wpr}$ that all the states reachable from $(s, (Y, B, L))$ have priority 2 according to $\widehat{p}$. Therefore ensuring that all recurrent classes reachable from $\widehat{S}_{wpr}$ have minimal priority 2. Due to the construction of the POMDP $\widehat{G}$, the only actions allowed in a state $(s, (Y, B, L))$ satisfy that for all states $\widehat{s} \in Y$ if $B(\widehat{s}) = 1$, $L(\widehat{s}) = \{Z_\infty\}$ and $\widehat{p}(s) \in Z_\infty$ for some $Z_\infty \subseteq \{1, 2\}$, then for all states $\widehat{s}' \in \mathsf{Supp}(\delta(s, a))$ we have that $p(\widehat{s}') \in Z_\infty$. As all states in $(s, (Y, B, L)) \in \widehat{S}_{wpr}$ have $L(s) = \{\{2\}\}$, it follows that any state reachable in the next step has priority 2. Let $(s', Y', (Y, B, L), a)$ be an arbitrary state reachable from $(s, (Y, B, L))$ in one step. By the previous argument we have that the priority $\widehat{p}((s', Y', (Y, B, L), a)) = 2$. Similarly the only allowed memory-update actions (Y', B', L') from state $(s', Y', (Y, B, L), a)$ satisfy that whenever $\widehat{s} \in Y$ and $B(\widehat{s}) = 1$, then for all $\widehat{s}' \in \mathsf{Supp}(\delta(\widehat{s}, a))$, we have that $B'(\widehat{s}') = 1$ and similarly we have that $L'(s')$ is a non-empty subset of $L(s)$, i.e., $L'(s') = \{\{2\}\}$. Therefore the next reachable state $(s', (Y', B', L'))$ is again in $\widehat{S}_{wpr}$. In other words, from states $(s, (Y, B, L))$ in $\widehat{S}_{wpr}$ in all future steps only states with priority 2 are visited, i.e., $\mathsf{Safe}(\widehat{p}^{-1}(2))$ is ensured which ensures the coBüchi objective. As the states in $\widehat{S}_{wpr}$ are reached with probability 1 and from them all recurrent classes reachable have only states that have priority 2, the desired result follows. $\square$

Lemma 25. $\mathsf{Almost}(\mathsf{coBuchi}(\widehat{p}^{-1}(2))) \subseteq W_2$.

Proof. Assume towards contradiction that there is an observation o in $\widehat{\mathcal{O}} \setminus W_2$ such that $o \in$ Almost(coBuchi($\widehat{p}^{-1}(2)$)). The observation o belongs to Almost(Safe(F)) as there is no winning strategy from observations outside Almost(Safe(F)), where $F = \widehat{S} \setminus \widehat{s}_b$. Consider a randomized memoryless strategy $\widehat{\sigma}$ satisfying the coBüchi objective from all observations in Almost(coBuchi($\widehat{p}^{-1}(2)$)). By Lemma 22 there exists a finite-memory almost-sure winning strategy σ in the POMDP G. Let us consider the almost-sure winning projected strategy $\sigma' = proj(\sigma)$ in the POMDP G. Recall that by Lemma 9 the set of pseudo-recurrent states is reached with probability 1 in the Markov chain $G \upharpoonright \sigma'$. By Lemma 12, for a pseudo-recurrent state $(s, (Y, B, L))$ there exists reachable recurrent class with the priority set as $L(s)$, and since σ' is an almost-sure winning strategy every recurrent class must have only priority 2, and hence for every reachable pseudo-recurrent state $(s, (Y, B, L))$ we must have $L(s) = \{\{2\}\}$. And by the definition of pseudo-recurrent states we also have that $B(s) = 1$ and $p(s) \in \{2\}$. As $\widehat{p}(s, (Y, B, L)) = p(s)$ we have that $(s, (Y, B, L)) \in \widehat{S}_{wpr}$. This implies that every pseudo-recurrent state reachable is a state in $\widehat{S}_{wpr}$. We want to show that in the construction described in Lemma 21, the memoryless almost-sure winning strategy $\widehat{\sigma}'$ constructed from the projected strategy σ' will ensure reaching the set $\widehat{S}_{wpr}$ in the Markov chain $\widehat{G} \upharpoonright \widehat{\sigma}'$ with probability 1. In the proof of Lemma 21 we have already established that reachability is preserved, i.e., if $(s', (Y', B', L'))$ is reachable from $(s, (Y, B, L))$ in $G \upharpoonright \sigma'$ then $(s', (Y', B', L'))$ is reachable from $(s, (Y, B, L))$ in $\widehat{G} \upharpoonright \widehat{\sigma}'$. As by Lemma 9 from every state a pseudo-recurrent state is reached with positive probability, and (as argued above every reachable pseudo-recurrent state is in $\widehat{S}_{wpr}$) we have that from every state in $\widehat{G} \upharpoonright \widehat{\sigma}'$ a state in $\widehat{S}_{wpr}$ is reachable. As this is true for every state we have that the set of states $\widehat{S}_{wpr}$ is reached with probability 1 in $\widehat{G} \upharpoonright \widehat{\sigma}'$ (Property 1 (a)). Therefore we have that the observation o belongs to Almost(Reach($\widehat{S}_{wpr}$)). But this contradicts that o does not belong to W_2 and the desired result follows. $\qquad\square$

To complete the computation for almost-sure winning for coBüchi objectives we now present polynomial time solutions for almost-sure safety and almost-sure Büchi objectives (that implies the solution for almost-sure reachability) in belief-observation POMDPs for randomized memoryless strategies. We start with a few notations below:

- *(Pre).* The predecessor function given a set of observations O selects the observations $o \in O$ such that Allow(o, O) is non-empty , i.e.,

$$\text{Pre}(O) = \{o \in O \mid \text{Allow}(o, O) \neq \emptyset\}.$$

- *(Apre).* Given a set $Y \subseteq \mathcal{O}$ of observations and a set $X \subseteq S$ of states such that $X \subseteq \gamma^{-1}(Y)$, the set Apre($Y, X$) denotes the states from $\gamma^{-1}(Y)$ such that there exists an action that ensures that the next observation is in Y and the set X is reached with positive probability, i.e.,:

$$\text{Apre}(Y, X) = \{s \in \gamma^{-1}(Y) \mid \exists a \in \text{Allow}(\gamma(s), Y) \text{ such that } \text{Supp}(\delta(s, a)) \cap X \neq \emptyset\}.$$

- *(ObsCover).* For a set $U \subseteq S$ of states we define the ObsCover(U) $\subseteq \mathcal{O}$ to be the set of observations o such that all states with observation o is in U, i.e., ObsCover(U) $= \{o \in \mathcal{O} \mid \gamma^{-1}(o) \subseteq U\}$.

Using the above notations we present the solution of almost-sure winning for safety and Büchi objectives.

Almost-sure winning for safety objectives. Given a safety objective Safe(F), for a set $F \subseteq S$ of states, let $O_F = \text{ObsCover}(F)$ denote the set of observations o such that $\gamma^{-1}(o) \subseteq F$, i.e., for all states $s \in \gamma^{-1}(o)$ belong to F. We denote by νX the greatest fixpoint and by μX the least fixpoint. Let

$$Y^* = \nu Y.(O_F \cap \text{Pre}(Y)) = \nu Y.(\text{ObsCover}(F) \cap \text{Pre}(Y))$$

be the greatest fixpoint of the function $f(Y) = O_F \cap \mathsf{Pre}(Y)$. Then the set Y^* is obtained by the following computation:

1. $Y_0 \leftarrow O_F$; and

2. repeat $Y_{i+1} \leftarrow \mathsf{Pre}(Y_i)$ until a fixpoint is reached.

We show that $Y^* = \mathsf{Almost}(\mathsf{Safe}(F))$.

Lemma 26. *For every observation $o \in Y^*$ we have $\mathit{Allow}(o, Y^*) \neq \emptyset$ (i.e., $\mathit{Allow}(o, Y^*)$ is non-empty).*

Proof. Assume towards contradiction that there exists an observation $o \in Y^*$ such that $\mathsf{Allow}(o, Y^*)$ is empty. Then $o \notin \mathsf{Pre}(Y^*)$ and hence the observation must be removed in the next iteration of the algorithm. This implies $\mathsf{Pre}(Y^*) \neq Y^*$, we reach a contradiction that Y^* is a fixpoint. $\square$

Lemma 27. *The set Y^* is the set of almost-sure winning observations for the safety objective $\mathsf{Safe}(F)$, i.e., $Y^* = \mathsf{Almost}(\mathsf{Safe}(F))$, and can be computed in linear time.*

Proof. We prove the two desired inclusions: (1) $Y^* \subseteq \mathsf{Almost}(\mathsf{Safe}(F))$; and (2) $\mathsf{Almost}(\mathsf{Safe}(F)) \subseteq Y^*$.

1. *(First inclusion).* By the definition of Y_0 we have that $\gamma^{-1}(Y_0) \subseteq F$. As $Y_{i+1} \subseteq Y_i$ we have that $\gamma^{-1}(Y^*) \subseteq F$. By Lemma 26, for all observations $o \in Y^*$ we have $\mathsf{Allow}(o, Y^*)$ is non-empty. A pure memoryless that plays some action from $\mathsf{Allow}(o, Y^*)$ in o, for $o \in Y^*$, ensures that the next observation is in Y^*. Thus the strategy ensures that only states from $\gamma^{-1}(Y^*) \subseteq F$ are visited, and therefore is an almost-sure winning strategy for the safety objective.

2. *(Second inclusion).* We prove that there is no almost-sure winning strategy from $\mathcal{O} \setminus Y^*$ by induction:

   - **(Base case).** There is no almost-sure winning strategy from observations $\mathcal{O} \setminus Y_0$. Note that $Y_0 = O_F$. In every observation $o \in \mathcal{O} \setminus Y_0$ there exists a state $s \in \gamma^{-1}(o)$ such that $s \notin F$. As G is a belief-observation POMDP there is a positive probability of being in state s, and therefore not being in F.

   - **(Inductive step).** We show that there is no almost-sure winning strategy from observations in $\mathcal{O} \setminus Y_{i+1}$. Let $Y_{i+1} \neq Y_i$ and $o \in Y_i \setminus Y_{i+1}$ (or equivalently $(\mathcal{O} \setminus Y_{i+1}) \setminus (\mathcal{O} \setminus Y_i)$). As the observation o is removed from Y_i it follows that $\mathsf{Allow}(o, Y_i) = \emptyset$. It follows that no matter what action is played, there is a positive probability of being in a state $s \in \gamma^{-1}(o)$ such that playing the action would leave the set $\gamma^{-1}(Y_i)$ with positive probability, and thus reaching the observations $\mathcal{O} \setminus Y_i$ from which there is no almost-sure winning strategy by induction hypothesis.

This shows that $Y^* = \mathsf{Almost}(\mathsf{Safe}(F))$, and the linear time computation follows from the straight forward computation of greatest fixpoints. The desired result follows. $\square$

We now present one simple lemma that was implicitly used in the restriction of the POMDP $\widehat{G}$ to almost-sure safety that a randomized memoryless strategies must only play action in the Allow set.

Lemma 28. *Let σ be a randomized memoryless almost-sure winning strategy in a belief-observation POMDP G for the safety objective $\mathsf{Safe}(F)$. Then $\mathrm{Supp}(\sigma(o)) \subseteq \mathit{Allow}(o, \mathsf{Almost}(\mathsf{Safe}(F)))$.*

Proof. Assume that the strategy σ plays an action $a \in A \setminus \mathsf{Allow}(o, \mathsf{Almost}(\mathsf{Safe}(F)))$ after an observation o. Then there is a positive probability of being in a state $s \in \gamma^{-1}(o)$ such that playing the action a in that state would leave the observations $\mathsf{Almost}(\mathsf{Safe}(F))$ with positive probability. As there is no randomized almost-sure winning strategy in $S \setminus \mathsf{Almost}(\mathsf{Safe}(F))$ (by definition), this contradicts the fact that σ is almost-sure winning. $\qquad\square$

Almost-sure winning for Büchi objectives. Consider a set $T \subseteq S$ of target states, and the Büchi objective $\mathsf{Buchi}(T)$. We will show that:

$$\mathsf{Almost}(\mathsf{Buchi}(T)) = \nu Z.\mathsf{ObsCover}(\mu X.((T \cap \gamma^{-1}(Z) \cap \gamma^{-1}(\mathsf{Pre}(Z))) \cup \mathsf{Apre}(Z, X))).$$

Let $Z^* = \nu Z.\mathsf{ObsCover}(\mu X.((T \cap \gamma^{-1}(Z) \cap \gamma^{-1}(\mathsf{Pre}(Z))) \cup \mathsf{Apre}(Z, X)))$. In the following two lemmas we show the two desired inclusions, i.e., $\mathsf{Almost}(\mathsf{Buchi}(T)) \subseteq Z^*$ and then we show that $Z^* \subseteq \mathsf{Almost}(\mathsf{Buchi}(T))$.

Lemma 29. $\mathsf{Almost}(\mathsf{Buchi}(T)) \subseteq Z^*$.

Proof. Let $W^* = \mathsf{Almost}(\mathsf{Buchi}(T))$. We first show that W^* is a fixpoint of the function

$$f(Z) = \mathsf{ObsCover}(\mu X.((T \cap \gamma^{-1}(Z) \cap \gamma^{-1}(\mathsf{Pre}(Z))) \cup \mathsf{Apre}(Z, X))),$$

i.e., we will show that $W^* = \mathsf{ObsCover}(\mu X.((T \cap \gamma^{-1}(W^*) \cap \gamma^{-1}(\mathsf{Pre}(W^*))) \cup \mathsf{Apre}(W^*, X)))$. As Z^* is the greatest fixpoint it will follow that $W^* \subseteq Z^*$.

Let

$$X^* = (\mu X.((T \cap \gamma^{-1}(W^*) \cap \gamma^{-1}(\mathsf{Pre}(W^*))) \cup \mathsf{Apre}(W^*, X))),$$

and $\widehat{X}^* = \mathsf{ObsCover}(X^*)$. Note that by definition we have $X^* \subseteq \gamma^{-1}(W^*)$ as the inner fixpoint computation only computes states that belong to $\gamma^{-1}(W^*)$. Assume towards contradiction that W^* is not a fixpoint, i.e., $\widehat{X}^*$ is a strict subset of W^*. For all states $s \in \gamma^{-1}(W^*) \setminus X^*$, for all actions $a \in \mathsf{Allow}(\gamma(s), W^*)$ we have $\mathsf{Supp}(\delta(s, a)) \subseteq (\gamma^{-1}(W^*) \setminus X^*)$. Consider any randomized memoryless almost-sure winning strategy σ^* from W^* and we consider two cases:

1. Suppose there is a state $s \in \gamma^{-1}(W^*) \setminus X^*$ such that an action that does not belong to $\mathsf{Allow}(\gamma(s), W^*)$ is played with positive probability by σ^*. Then with positive probability the observations from W^* are left (because from some state with same observation as s an observation in the complement of W^* is reached with positive probability). Since from the complement of W^* there is no randomized memoryless almost-sure winning strategy (by definition), it contradicts that σ^* is an almost-sure winning strategy from W^*.

2. Otherwise for all states $s \in \gamma^{-1}(W^*) \setminus X^*$ the strategy σ^* plays only actions in $\mathsf{Allow}(\gamma(s), W^*)$, and then the probability to reach X^* is zero, i.e., $\mathsf{Safe}(\gamma^{-1}(W^*) \setminus X^*)$ is ensured. Since all target states in $\gamma^{-1}(W^*)$ belong to X^* (they get included in iteration 0 of the fixpoint computation) it follows that $(\gamma^{-1}(W^*) \setminus X^*) \cap T = \emptyset$, and hence $\mathsf{Safe}(\gamma^{-1}(W^*) \setminus X^*) \cap \mathsf{Buchi}(T) = \emptyset$, and we again reach a contradiction that σ^* is an almost-sure winning strategy.

It follows that W^* is a fixpoint, and thus we get that $W^* \subseteq Z^*$. $\qquad\square$

Lemma 30. $Z^* \subseteq \mathsf{Almost}(\mathsf{Buchi}(T))$.

Proof. We define a randomized memoryless strategy σ^* for the objective $\mathsf{Almost}(\mathsf{Buchi}(T))$ as follows: for an observation $o \in Z^*$, play all actions from the set $\mathsf{Allow}(o, Z^*)$ uniformly at random. Since the strategy σ^* plays only actions in $\mathsf{Allow}(o, Z^*)$, for $o \in Z^*$, it ensures that the set of states $\gamma^{-1}(Z^*)$ is not left, (i.e., $\mathsf{Safe}(\gamma^{-1}(Z^*))$ is ensured). We now analyze the computation of the inner fixpoint, i.e., analyze the computation of $\mu X.((T \cap \gamma^{-1}(Z^*) \cap \gamma^{-1}(\mathsf{Pre}(Z^*))) \cup \mathsf{Apre}(Z^*, X)))$ as follows:

- $X_0 = (T \cap \gamma^{-1}(Z^*) \cap \gamma^{-1}(\mathsf{Pre}(Z^*))) \cup \mathsf{Apre}(Z^*, \emptyset))) = T \cap \gamma^{-1}(Z^*) \cap \gamma^{-1}(\mathsf{Pre}(Z^*)) \subseteq T$ (since $\mathsf{Apre}(Z^*, \emptyset)$ is emptyset);

- $X_{i+1} = (T \cap \gamma^{-1}(Z^*) \cap \gamma^{-1}(\mathsf{Pre}(Z^*))) \cup \mathsf{Apre}(Z^*, X_i)))$

Note that we have $X_0 \subseteq T$. For every state $s_j \in X_j$ the set of played actions $\mathsf{Allow}(\gamma(s_j), Z^*)$ contains an action a such that $\mathsf{Supp}(\delta(s_j, a)) \cap X_{j-1}$ is non-empty. Let C be an arbitrary reachable recurrent class in the Markov Chain $G \restriction \sigma$ reachable from a state in $\gamma^{-1}(Z^*)$. Since $\mathsf{Safe}(\gamma^{-1}(Z^*))$ is ensured, it follows that $C \subseteq \gamma^{-1}(Z^*)$. Consider a state in C that belongs to $X_j \setminus X_{j-1}$ for $j \geq 1$. Since the strategy ensures that for some action a played with positive probability we must have $\mathsf{Supp}(\delta(s_j, a)) \cap X_{j-1} \neq \emptyset$, it follows that $C \cap X_{j-1} \neq \emptyset$. Hence by induction $C \cap X_0 \neq \emptyset$. It follows $C \cap T \neq \emptyset$. Hence all reachable recurrent classes intersect with the target set and thus the strategy σ^* ensures that T is visited infinitely often with probability 1. Thus we have $Z^* \subseteq \mathsf{Almost}(\mathsf{Buchi}(T))$. □

Lemma 31. *The set* $\mathsf{Almost}(\mathsf{Buchi}(T))$ *and* $\mathsf{Almost}(\mathsf{Reach}(T))$ *can be computed in quadratic time for belief-observation POMDPs, for target set* $T \subseteq S$.

Proof. For $\mathsf{Almost}(\mathsf{Buchi}(T))$ it follows directly from Lemma 29 and Lemma 30. The result for $\mathsf{Almost}(\mathsf{Reach}(T))$ follows from the fact that $\mathsf{Reach}(T)$ is a special case of $\mathsf{Buchi}(T)$ (by converting states in the target set T to absorbing states). □

The EXPTIME-completeness. In Section 4 we have established a polynomial time reduction of POMDPs with parity objectives to POMDPs with coBüchi objectives for almost-sure winning under finite-memory strategies. In this section we first showed that given a POMDP G with a coBüchi objective we can construct an exponential size belief-observation POMDP $\widehat{G}$ and the computation of the almost-sure winning set for coBüchi objectives reduced to the computation of the almost-sure winning set for safety and reachability objectives, for which we established linear and quadratic time algorithms respectively. This gives us an $2^{O(|S| \cdot d)}$ time algorithm to decide (and construct if one exists) the existence of finite-memory almost-sure winning strategies in POMDPs with parity objectives with d priorities. The EXPTIME-hardness follows from the results of [8] that shows deciding the existence of finite-memory almost-sure winning strategies in POMDPs with reachability objectives is EXPTIME-hard. The results for positive winning goes via reduction to Büchi objectives and is similar. We have the following result.

Theorem 4. *The following assertions hold:*

1. *Given a POMDP G with $|S|$ states and a parity objective with d priorities, the existence (and the construction if one exists) of a finite-memory almost-sure (resp. positive) winning strategy can be achieved in $2^{O(|S| \cdot d)}$ time.*

2. *The decision problem of given a POMDP and a parity objective whether there exists a finite-memory almost-sure (resp. positive) winning strategy is EXPTIME-complete.*

Remark 3. *Note that our EXPTIME-algorithm for parity objectives, and the LAR reduction of Muller objectives to parity objectives [18] give an $2^{O(d! \cdot d^2 \cdot |S|)}$ time algorithm for Muller objectives with d colors for POMDPs with $|S|$ states, i.e., the algorithm is exponential in $|S|$ and double exponential in d. Note that the Muller objective specified by the set $\mathcal{F}$ maybe in general itself double exponential in d.*

References

[1] C. Baier, N. Bertrand, and M. Größer. On decision problems for probabilistic Büchi automata. In *FoSSaCS*, LNCS 4962, pages 287–301. Springer, 2008.

[2] C. Baier, M. Größer, and N. Bertrand. Probabilistic omega-automata. *J. ACM*, 59(1), 2012.

[3] C. Baier and J-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.

[4] A. Bianco and L. de Alfaro. Model checking of probabilistic and nondeterministic systems. In *FSTTCS 95*, volume 1026 of *LNCS*, pages 499–513. Springer-Verlag, 1995.

[5] P. Billingsley, editor. *Probability and Measure*. Wiley-Interscience, 1995.

[6] P. Cerný, K. Chatterjee, T. A. Henzinger, A. Radhakrishna, and R. Singh. Quantitative synthesis for concurrent programs. In *Proc. of CAV*, LNCS 6806, pages 243–259. Springer, 2011.

[7] K. Chatterjee, L. Doyen, H. Gimbert, and T. A. Henzinger. Randomness for free. In *MFCS*, pages 246–257, 2010.

[8] K. Chatterjee, L. Doyen, and T. A. Henzinger. Qualitative analysis of partially-observable Markov decision processes. In *MFCS*, pages 258–269, 2010.

[9] K. Chatterjee, L. Doyen, T.A. Henzinger, and J.F. Raskin. Algorithms for omega-regular games with imperfect information. In *CSL'06*, pages 287–302. LNCS 4207, Springer, 2006.

[10] K. Chatterjee and T. A. Henzinger. Probabilistic automata on infinite words: Decidability and unde-cidability results. In *ATVA*, pages 1–16, 2010.

[11] A. Condon and R. J. Lipton. On the complexity of space bounded interactive proofs. In *FOCS*, pages 462–467, 1989.

[12] C. Courcoubetis and M. Yannakakis. The complexity of probabilistic verification. *Journal of the ACM*, 42(4):857–907, 1995.

[13] K. Culik and J. Kari. Digital images and formal languages. *Handbook of formal languages*, pages 599–616, 1997.

[14] L. de Alfaro, M. Faella, R. Majumdar, and V. Raman. Code-aware resource management. In *EMSOFT 05*. ACM, 2005.

[15] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge Univ. Press, 1998.

[16] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer-Verlag, 1997.

[17] H. Gimbert and Y. Oualhadj. Probabilistic automata on finite words: Decidable and undecidable problems. In *Proc. of ICALP*, LNCS 6199, pages 527–538. Springer, 2010.

[18] Y. Gurevich and L. Harrington. Trees, automata, and games. In *STOC'82*, pages 60–65. ACM Press, 1982.

[19] H. Howard. *Dynamic Programming and Markov Processes*. MIT Press, 1960.

[20] L.P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *J. of Artif. Intell. Research*, 4:237–285, 1996.

[21] A. Kechris. *Classical Descriptive Set Theory*. Springer, 1995.

[22] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas. Temporal-logic-based reactive mission and motion planning. *IEEE Transactions on Robotics*, 25(6):1370–1381, 2009.

[23] M. Kwiatkowska, G. Norman, and D. Parker. Verifying randomized distributed algorithms with prism. In *Workshop on Advances in Verification (WAVE'00)*, 2000.

[24] M. Kwiatkowska, G. Norman, and D. Parker. PRISM: Probabilistic symbolic model checker. In *TOOLS' 02*, pages 200–204. LNCS 2324, Springer, 2002.

[25] O. Madani, S. Hanks, and A. Condon. On the undecidability of probabilistic planning and related stochastic optimization problems. *Artif. Intell.*, 147(1-2):5–34, 2003.

[26] M. Mohri. Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(2):269–311, 1997.

[27] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of Markov decision processes. *Mathematics of Operations Research*, 12:441–450, 1987.

[28] A. Paz. *Introduction to probabilistic automata (Computer science and applied mathematics)*. Academic Press, 1971.

[29] A. Pogosyants, R. Segala, and N. Lynch. Verification of the randomized consensus algorithm of Aspnes and Herlihy: a case study. *Distributed Computing*, 13(3):155–186, 2000.

[30] M.O. Rabin. Probabilistic automata. *Information and Control*, 6:230–245, 1963.

[31] J. H. Reif. Universal games of incomplete information. In *STOC*, pages 288–308. ACM Press, 1979.

[32] J. H. Reif. The complexity of two-player games of incomplete information. *Journal of Computer and System Sciences*, 29(2):274–301, 1984.

[33] R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995. Technical Report MIT/LCS/TR-676.

[34] M.I.A. Stoelinga. Fun with FireWire: Experiments with verifying the IEEE1394 root contention protocol. In *Formal Aspects of Computing*, 2002.

[35] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.