Energy and Mean-Payoff Parity Markov Decision Processes ⋆

Krishnendu Chatterjee[1] and Laurent Doyen[2]

[1] IST Austria (Institute of Science and Technology Austria)
[2] LSV, ENS Cachan & CNRS, France

Abstract. We consider Markov Decision Processes (MDPs) with mean-payoff parity and energy parity objectives. In system design, the parity objective is used to encode ω-regular specifications, while the mean-payoff and energy objectives can be used to model quantitative resource constraints. The energy condition requires that the resource level never drops below 0, and the mean-payoff condition requires that the limit-average value of the resource consumption is within a threshold. While these two (energy and mean-payoff) classical conditions are equivalent for two-player games, we show that they differ for MDPs. We show that the problem of deciding whether a state is almost-sure winning (i.e., winning with probability 1) in energy parity MDPs is in NP $\cap$ coNP, while for mean-payoff parity MDPs, the problem is solvable in polynomial time.

1 Introduction

Markov decision processes (MDPs) are a standard model for systems that exhibit both stochastic and nondeterministic behaviour. The nondeterminism represents the freedom of choice of control actions, while the probabilities describe the uncertainty in the response of the system to control actions. The control problem for MDPs asks whether there exists a strategy (or policy) to select control actions in order to achieve a given goal with a certain probability. MDPs have been used in several areas such as planning, probabilistic reactive programs, verification and synthesis of (concurrent) probabilistic systems [12, 22, 1].

The control problem may specify a goal as a set of desired traces (such as ω-regular specifications), or as a quantitative optimization objective for a payoff function defined on the traces of the MDP. Typically, discounted-payoff and mean-payoff functions have been studied [15]. Recently, the energy objectives (corresponding to total-payoff functions) have been considered in the design of resource-constrained embedded systems [3, 7, 20] such as power-limited systems, as well as in queueing processes, and gambling models (see also [4] and references therein). The energy objective requires that the sum of the rewards be always nonnegative along a trace. Energy objective can be expressed in the setting of boundaryless one-counter MDPs [4]. In the case of MDPs, achieving energy objective with probability 1 is equivalent to achieving energy objective in the stronger setting of a two-player game where the probabilistic choices are replaced by

⋆ This work was partially supported by FWF NFN Grant S11407-N23 (RiSE) and a Microsoft faculty fellowship.

adversarial choice. This is because if a trace ρ violates the energy condition in the game, then a finite prefix of ρ would have a negative energy, and this finite prefix has positive probability in the MDP. Note that in the case of two-player games, the energy objective is equivalent to enforce nonnegative mean-payoff value [3, 5].

In this paper, we consider MDPs equipped with the combination of a parity objective (which is a canonical way to express the ω-regular conditions [21]), and a quantitative objective specified as either mean-payoff or energy condition. Special cases of the parity objective include reachability and fairness objectives such as Büchi and coBüchi conditions. Such combination of quantitative and qualitative objectives is crucial in the design of reactive systems with both resource constraints and functional requirements [6, 11, 3, 2]. For example, Kucera and Stražovský consider the combination of PCTL with mean-payoff objectives for MDPs and present an EXPTIME algorithm [19]. In the case of energy parity condition, it can also be viewed as a natural extension of boundaryless one-counter MDPs with fairness conditions.

Consider the MDP in Fig. 1, with the objective to visit the Büchi state q_2 infinitely often, while maintaining the energy level (i.e., the sum of the transition weights) nonnegative. A winning strategy from q_0 would loop 20 times on q_0 to accumulate energy and then it can afford to reach the probabilistic state from which the Büchi state is reached with probability $1/2$ and cost 20. If the Büchi state is not reached immediately, then the strategy needs to recharge 10 units of energy and try again. This strategy uses memory and it is also winning with probability 1 for the nonnegative mean-payoff Büchi objective. In general however, the energy and mean-payoff parity objectives do not coincide (see later the example in Fig. 2). In particular, the memory requirement for energy parity objective is finite (at most exponential) while it may be infinite for mean-payoff parity.

We study the computational complexity of the problem of deciding if there exists a strategy to achieve energy parity objective, or mean-payoff parity objective with probability 1 (i.e., almost-surely). We provide the following bounds for these problems.

1. For energy parity MDPs, we show that the problem is in NP $\cap$ coNP, and present a pseudo-polynomial time algorithm. Since parity games polynomially reduce to two-player energy games [18, 3, 5], and thus to energy MDPs, the problem for almost-sure energy parity MDPs is at least as hard as solving two-player parity games.
2. For mean-payoff parity MDPs, we show that the problem is solvable in polynomial time (and thus PTIME-complete).

We refer to [12, 16, 9] for importance of the computation of almost-sure winning set related to robust solutions (independence of precise transition probabilities) and the more general quantitative problem. The computation of the almost-sure winning set in MDPs typically relies either on the end-component analysis, or analysis of attractors and sub-MDPs. Our results for mean-payoff parity objectives rely on the end-component analysis, but in a more refined way than the standard analysis, to obtain a polynomial-time algorithm. Our proof combines techniques for mean-payoff and parity objectives to produce infinite-memory strategy witnesses, which is necessary in general. We present an algorithm that iterates successively over even priorities $2i$ and computes almost-sure winning end-components with the even priority $2i$ as the best priority. The problem

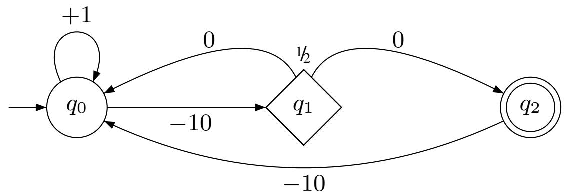

Fig. 1. An energy Büchi MDP. The player-1 states are q_0, q_2, and the probabilistic state is q_1.

of positive mean-payoff objectives and parity objectives has been considered independently in [17].

For energy parity MDPs the end-component based analysis towards polynomial-time algorithm does not work since solving energy parity MDPs is at least as hard as solving two-player parity games. Instead, for energy parity MDPs, we present a quadratic reduction to two-player energy Büchi games which are in NP $\cap$ coNP and solvable in pseudo-polynomial time [7].

From our results, it follows that for energy parity MDPs, strategies with finite memory are sufficient (linear in the number of states times the value of the largest weight), while infinite memory may be necessary for mean-payoff parity MDPs. The details of the proofs can be found in [8], as well as the solution for disjunction of mean-payoff parity and energy parity objectives. An interesting open question is to extend the results of this paper from MDPs to two-player stochastic games.

2 Definitions

Probability distributions. A *probability distribution* over a finite set A is a function $\kappa : A \to [0,1]$ such that $\sum_{a \in A} \kappa(a) = 1$. The *support* of κ is the set $\mathsf{Supp}(\kappa) = \{a \in A \mid \kappa(a) > 0\}$. We denote by $\mathcal{D}(A)$ the set of probability distributions on A.

Markov Decision Processes. A *Markov Decision Process* (MDP) $M = (Q, E, \delta)$ consists of a finite set Q of states partitioned into *player-*1 *states* Q_1 and *probabilistic states* Q_P (i.e., $Q = Q_1 \cup Q_P$ and $Q_1 \cap Q_P = \emptyset$), a set $E \subseteq Q \times Q$ of edges such that for all $q \in Q$, there exists (at least one) $q' \in Q$ such that $(q, q') \in E$, and a probabilistic transition function $\delta : Q_P \to \mathcal{D}(Q)$ such that for all $q \in Q_P$ and $q' \in Q$, we have $(q, q') \in E$ iff $\delta(q)(q') > 0$. We often write $\delta(q, q')$ for $\delta(q)(q')$. For a state $q \in Q$, we denote by $E(q) = \{q' \in Q \mid (q, q') \in E\}$ the set of possible successors of q.

End-components and Markov chains. A set $U \subseteq Q$ is δ*-closed* if for all $q \in U \cap Q_P$ we have $\mathsf{Supp}(\delta(q)) \subseteq U$. The sub-MDP induced by a δ-closed set U is $M \upharpoonright U = (U, E \cap (U \times U), \delta)$. Note that $M \upharpoonright U$ is an MDP if for all $q \in U$ there exists $q' \in U$ such that $(q, q') \in E$. A *Markov chain* is a special case of MDP where $Q_1 = \emptyset$. A *closed recurrent set* for a Markov chain is a δ-closed set $U \subseteq Q$ which is strongly connected. End-components in MDPs play a role equivalent to closed recurrent sets in Markov chains. Given an MDP $M = (Q, E, \delta)$ with partition (Q_1, Q_P), a set $U \subseteq Q$

of states is an *end-component* if U is δ-closed and the sub-MDP $M \restriction U$ is strongly connected [12, 13]. We denote by $\mathcal{E}(M)$ the set of end-components of an MDP M.

Plays. An MDP can be viewed as the arena of a game played for infinitely many rounds from a state $q_0 \in Q$ as follows. If the game is in a player-1 state q, then player 1 chooses the successor state in the set $E(q)$; otherwise the game is in a probabilistic state q, and the successor is chosen according to the probability distribution $\delta(q)$. This game results in a *play* from q_0, i.e., an infinite path $\rho = q_0 q_1 \ldots$ such that $(q_i, q_{i+1}) \in E$ for all $i \geq 0$. The prefix of length n of ρ is denoted by $\rho(n) = q_0 \ldots q_n$, the last state of $\rho(n)$ is $\mathsf{Last}(\rho(n)) = q_n$. We write Ω for the set of all plays.

Strategies. A *strategy* (for player 1) is a function $\sigma : Q^* Q_1 \to \mathcal{D}(Q)$ such that for all $\rho \in Q^*$, $q \in Q_1$, and $q' \in Q$, if $\sigma(\rho \cdot q)(q') > 0$, then $(q, q') \in E$. We denote by Σ the set of all strategies. An *outcome* of σ from q_0 is a play $q_0 q_1 \ldots$ where $q_{i+1} \in \mathsf{Supp}(\sigma(q_0 \ldots q_i))$ for all $i \geq 0$ such that $q_i \in Q_1$. Strategies that do not use randomization are called pure. A player-1 strategy σ is *pure* if for all $\rho \in Q^*$ and $q \in Q_1$, there is a state $q' \in Q$ such that $\sigma(\rho \cdot q)(q') = 1$.

Outcomes and measures. Once a starting state $q \in Q$ and a strategy $\sigma \in \Sigma$ are fixed, the outcome of the game is a random walk ω_q^σ for which the probabilities of every *event* $\mathcal{A} \subseteq \Omega$, which is a measurable set of plays, are uniquely defined [22]. For a state $q \in Q$ and an event $\mathcal{A} \subseteq \Omega$, we denote by $\mathbb{P}_q^\sigma(\mathcal{A})$ the probability that a play belongs to $\mathcal{A}$ if the game starts from the state q and player 1 follows the strategy σ. For a measurable function $f : \Omega \to \mathbb{R}$ we denote by $\mathbb{E}_q^\sigma[f]$ the *expectation* of the function f under the probability measure $\mathbb{P}_q^\sigma(\cdot)$.

Finite-memory strategies. A strategy uses *finite-memory* if it can be encoded by a deterministic transducer $\langle \mathsf{Mem}, m_0, \alpha_u, \alpha_n \rangle$ where Mem is a finite set (the memory of the strategy), $m_0 \in \mathsf{Mem}$ is the initial memory value, $\alpha_u : \mathsf{Mem} \times Q \to \mathsf{Mem}$ is an update function, and $\alpha_n : \mathsf{Mem} \times Q_1 \to \mathcal{D}(Q)$ is a next-move function. The *size* of the strategy is the number $|\mathsf{Mem}|$ of memory values. If the game is in a player-1 state q, and m is the current memory value, then the strategy chooses the next state q' according to the probability distribution $\alpha_n(m, q)$, and the memory is updated to $\alpha_u(m, q)$. Formally, $\langle \mathsf{Mem}, m_0, \alpha_u, \alpha_n \rangle$ defines the strategy σ such that $\sigma(\rho \cdot q) = \alpha_n(\hat{\alpha}_u(m_0, \rho), q)$ for all $\rho \in Q^*$ and $q \in Q_1$, where $\hat{\alpha}_u$ extends α_u to sequences of states as expected. A strategy is *memoryless* if $|\mathsf{Mem}| = 1$. For a finite-memory strategy σ, let M_σ be the Markov chain obtained as the product of M with the transducer defining σ, where $(\langle m, q \rangle, \langle m', q' \rangle)$ is an edge in M_σ if $m' = \alpha_u(m, q)$ and either $q \in Q_1$ and $q' \in \mathsf{Supp}(\alpha_n(m, q))$, or $q \in Q_P$ and $(q, q') \in E$.

Two-player games. A *two-player game* is a graph $G = (Q, E)$ with the same assumptions as for MDP, except that the partition of Q is denoted (Q_1, Q_2) where Q_2 is the set of *player-2 states*. The notions of play, strategies (in particular strategies for player 2), and outcome are analogous to the case of MDP [7].

Objectives. An *objective* for an MDP M (or game G) is a set $\phi \subseteq \Omega$ of infinite paths. Let $p : Q \to \mathbb{N}$ be a *priority function* and $w : E \to \mathbb{Z}$ be a *weight function* where positive numbers represent rewards. We denote by W the largest weight (in absolute value) according to w. The *energy level* of a prefix $\gamma = q_0 q_1 \ldots q_n$ of a play is

$\mathsf{EL}(w,\gamma) = \sum_{i=0}^{n-1} w(q_i, q_{i+1})$, and the *mean-payoff value*[3] of a play $\rho = q_0 q_1 \dots$ is $\mathsf{MP}(w,\rho) = \liminf_{n\to\infty} \frac{1}{n} \cdot \mathsf{EL}(w, \rho(n))$. In the sequel, when the weight function w is clear from the context we omit it and simply write $\mathsf{EL}(\gamma)$ and $\mathsf{MP}(\rho)$. We denote by $\mathsf{Inf}(\rho)$ the set of states that occur infinitely often in ρ, and we consider the following objectives:

- *Parity objectives.* The *parity* objective $\mathsf{Parity}(p) = \{\rho \in \Omega \mid \min\{p(q) \mid q \in \mathsf{Inf}(\rho)\}$ is even $\}$ requires that the minimum priority visited infinitely often be even. The special cases of *Büchi* and *coBüchi* objectives correspond to the case with two priorities, $p : Q \to \{0,1\}$ and $p : Q \to \{1,2\}$ respectively.
- *Energy objectives.* Given an initial credit $c_0 \in \mathbb{N}$, the *energy* objective $\mathsf{PosEnergy}(c_0) = \{\rho \in \Omega \mid \forall n \geq 0 : c_0 + \mathsf{EL}(\rho(n)) \geq 0\}$ requires that the energy level be always positive.
- *Mean-payoff objectives.* Given a threshold $\nu \in \mathbb{Q}$, the *mean-payoff* objective $\mathsf{MeanPayoff}^{\geq\nu} = \{\rho \in \Omega \mid \mathsf{MP}(\rho) \geq \nu\}$ (resp. $\mathsf{MeanPayoff}^{>\nu} = \{\rho \in \Omega \mid \mathsf{MP}(\rho) > \nu\}$) requires that the mean-payoff value be at least ν (resp. strictly greater than ν).
- *Combined objectives.* The *energy parity* objective $\mathsf{Parity}(p) \cap \mathsf{PosEnergy}(c_0)$ and the *mean-payoff parity* objective $\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{\sim\nu}$ (for $\sim \in \{\geq, >\}$) combine the requirements of parity and energy (resp., mean-payoff) objectives.

Almost-sure winning strategies. For MDPs, we say that a player-1 strategy σ is *almost-sure winning* in a state q for an objective ϕ if $\mathbb{P}_q^\sigma(\phi) = 1$. For two-player games, we say that a player-1 strategy σ is *winning* in a state q for an objective ϕ if all outcomes of σ starting in q belong to ϕ. For energy objectives with unspecified initial credit, we also say that a strategy is (almost-sure) winning if it is (almost-sure) winning for *some* finite initial credit.

Decision problems. We are interested in the following problems. Given an MDP M with weight function w and priority function p, and a state q_0,

- the *energy parity problem* asks whether there exists a finite initial credit $c_0 \in \mathbb{N}$ and an almost-sure winning strategy for the energy parity objective from q_0 with initial credit c_0. We are also interested in computing the *minimum initial credit* in q_0 which is the least value of initial credit for which there exists an almost-sure winning strategy for player 1 in q_0. A strategy for player 1 is *optimal* in q_0 if it is winning from q_0 with the minimum initial credit;
- the *mean-payoff parity problem* asks whether there exists an almost-sure winning strategy for the mean-payoff parity objective with threshold 0 from q_0. Note that it is not restrictive to consider mean-payoff objectives with threshold 0 because for $\sim \in \{\geq, >\}$, we have $\mathsf{MP}(w,\rho) \sim \nu$ iff $\mathsf{MP}(w - \nu, \rho) \sim 0$, where $w - \nu$ is the weight function that assigns $w(e) - \nu$ to each edge $e \in E$.

The two-player game version of these problems is defined analogously [7]. It is known that the initial credit problem for two-player energy games [6, 3], as well as two-player parity games [14] can be solved in NP $\cap$ coNP because memoryless strategies are sufficient to win. Moreover, parity games reduce in polynomial time to mean-payoff games [18], which are log-space equivalent to energy games [3, 5]. It is a long-standing

[3] The results of this paper hold for the definition of mean-payoff value using $\limsup$ instead of $\liminf$.

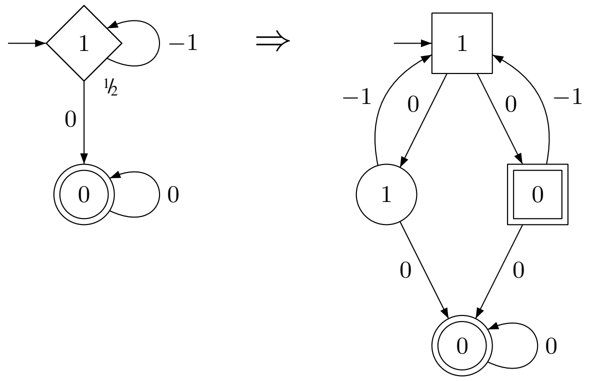

Fig. 2. The gadget construction is wrong for mean-payoff parity MDPs. Player 1 is almost-sure winning for mean-payoff Büchi in the MDP (on the left) but player 1 is losing in the two-player game (on the right) because player 2 (box-player) can force a negative-energy cycle.

open question to know if a polynomial-time algorithm exists for these problems. Finally, energy parity games and mean-payoff parity games are solvable in NP $\cap$ coNP although winning strategies may require exponential and infinite memory respectively, even in one-player games (and thus also in MDPs) [11, 7].

The decision problem for MDPs with parity objective, as well as with mean-payoff objective, can be solved in polynomial time [15, 12, 9, 13]. However, the problem is in NP $\cap$ coNP for MDPs with energy objective because an MDP with energy objective is equivalent to a two-player energy game (where the probabilistic states are controlled by player 2). Indeed (1) a winning strategy in the game is trivially almost-sure winning in the MDP, and (2) if an almost-sure winning strategy σ in the MDP was not winning in the game, then for all initial credit c_0 there would exist an outcome ρ of σ such that $c_0 + \mathsf{EL}(\rho(i)) < 0$ for some position $i \geq 0$. The prefix $\rho(i)$ has a positive probability in the MDP, in contradiction with the fact that σ is almost-sure winning. As a consequence, solving MDP with energy objectives is at least as hard as solving parity games.

In this paper, we show that the decision problem for MDPs with energy parity objective is in NP $\cap$ coNP, which is the best conceivable upper bound unless parity games can be solved in P. And for MDPs with mean-payoff parity objective, we show that the decision problem can be solved in polynomial time. The problem for MDPs with mean-payoff parity objectives under expectation semantics was considered in [10], whereas our semantics (threshold semantics) is different (we require the set of paths that satisfy the mean-payoff threshold has probability 1 rather than the expected value satisfy threshold).

The MDP in Fig. 2 on the left, which is essentially a Markov chain, is an example where the mean-payoff parity condition is satisfied almost-surely, while the energy parity condition is not, no matter the value of the initial credit. For initial credit c_0, the energy will drop below 0 with positive probability, namely $\frac{1}{2^{c_0+1}}$.

End-component lemma. We now present an important lemma about end-components from [12, 13] that we use in the proofs of our result. It states that for arbitrary strategies

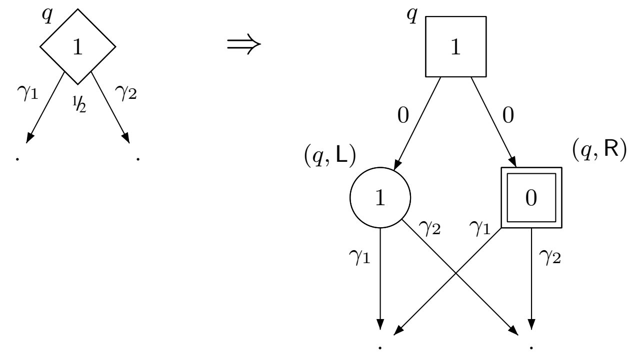

Fig. 3. Gadget for probabilistic states in energy Büchi MDP. Diamonds are probabilistic states, circles are player 1 states, and boxes are player 2 states.

(memoryless or not), with probability 1 the set of states visited infinitely often along a play is an end-component. This lemma allows us to derive conclusions on the (infinite) set of plays in an MDP by analyzing the (finite) set of end-components in the MDP.

Lemma 1. *[12, 13] Given an MDP M, for all states $q \in Q$ and all strategies $\sigma \in \Sigma$, we have* $\mathbb{P}_q^\sigma(\{\omega \mid \mathsf{Inf}(\omega) \in \mathcal{E}(M)\}) = 1$.

3 MDPs with Energy Parity Objectives

We show that energy parity MDPs can be solved in NP $\cap$ coNP, using a reduction to two-player energy Büchi games. Our reduction also preserves the value of the minimum initial credit. Therefore, we obtain a pseudo-polynomial algorithm for this problem, which also computes the minimum initial credit. Moreover, we show that the memory requirement for almost-sure winning strategies is at most $2{\cdot}|Q|{\cdot}W$, which is essentially optimal[4].

We first establish the results for the special case of energy Büchi MDPs. We present a reduction of the energy Büchi problem for MDPs to the energy Büchi problem for two-player games. The result then follows from the fact that the latter problem is in NP $\cap$ coNP and solvable in pseudo-polynomial time [7].

Given an MDP M, we can assume without loss of generality that every probabilistic state has priority 1, and has two outgoing transitions with probability ½ each [23, Section 6]. We construct a two-player game G by replacing every probabilistic state of M by a gadget as in Fig. 3. The probabilistic states q of M are mapped to player-2 states in G with two successors (q, L) and (q, R). Intuitively, player 2 chooses (q, L) to check whether player 1 can enforce the Büchi condition almost-surely. This is the case if player 1 can reach a Büchi state (with priority 0) infinitely often when he controls the probabilistic states (otherwise, no Büchi state is ever visited, and since $(\cdot, \mathsf{L})$ states have priority 1, the Büchi condition is not realized in G). And player 2 chooses (q, R)

[4] Example 1 in [7] shows that memory of size $2{\cdot}(|Q|-1){\cdot}W+1$ may be necessary.

to check that the energy condition is satisfied. If player 2 can exhaust the energy level in G, then the corresponding play prefix has positive probability in M. Note that (q, R) has priority 0 and thus cannot be used by player 2 to spoil the Büchi condition.

Formally, given $M = (Q, E, \delta)$ with partition (Q_1, Q_P) of Q, we construct a game $G = (Q', E')$ with partition (Q'_1, Q'_P) where $Q'_1 = Q_1 \cup (Q_P \times \{\mathsf{L}\})$ and $Q'_2 = Q_P \cup (Q_P \times \{\mathsf{R}\})$, see also Fig. 3. The states in Q' that are already in Q get the same priority as in M, the states $(\cdot, \mathsf{L})$ have priority 1, and the states $(\cdot, \mathsf{R})$ have priority 0. The set E' contains the following edges:

- all edges $(q, q') \in E$ such that $q \in Q_1$;
- edges $(q, (q, d))$, $((q, d), q')$ for all $q \in Q_P$, $d \in \{\mathsf{L}, \mathsf{R}\}$, and $q' \in \mathsf{Supp}(\delta(q))$.

The edges (q, q') and $((q, d), q')$ in E' get the same weight as (q, q') in M, and all edges $(q, (q, d))$ get weight 0.

Lemma 2. *Given an MDP M with energy Büchi objective, we can construct in linear time a two-player game G with energy Büchi objective such that for all states q_0 in M, there exists an almost-sure winning strategy from q_0 in M if and only if there exists a winning strategy from q_0 in G (with the same initial credit).*

Note that the reduction presented in the proof of Lemma 2 would not work for mean-payoff Büchi MDPs. Consider the MDP on Fig. 2 for which the gadget-based reduction to two-player games is shown on the right. The game is losing for player 1 both for energy and mean-payoff parity, simply because player 2 can always choose to loop through the box states, thus realizing a negative energy and mean-payoff value (no matter the initial credit). However player 1 is almost-sure winning in the mean-payoff parity MDP (on the left in Fig. 2).

While the reduction in the proof of Lemma 2 gives a game with $n' = |Q_1| + 3 \cdot |Q_P|$ states, the structure of the gadgets (see Fig. 3) is such that the energy level is independent of which of the transitions $(q, (q, \mathsf{L}))$ or $(q, (q, \mathsf{R}))$ is taken. Since from the result of [7, Lemma 8] and its proof, it follows that the memory updates in winning strategies for energy Büchi games can be done according to the energy level of the play prefix, it follows that the memory bound of $2 \cdot n \cdot W$ can be transfered to almost-sure winning strategies in Energy Büchi MDPs, where $n = |\mathsf{Win} \cap Q_1|$ is the number of player-1 almost-sure winning states. Also, the pseudo-polynomial algorithm for solving two-player energy Büchi games can be used for MDPs, with the same $O(|E| \cdot |Q|^5 \cdot W)$ complexity [7, Table 1] .

Using Lemma 2, we solve energy parity MDPs by a reduction to energy Büchi MDPs. The key idea of the reduction is that if player 1 has an almost-sure winning strategy for the energy parity objective, then player 1 can choose an even priority $2i$ and decide to satisfy the energy objective along with satisfying that priority $2i$ is visited infinitely often, and priorities less than $2i$ are visited finitely often.

W.l.o.g. we assume that player-1 states and probabilistic states alternate, i.e. $E(q) \subseteq Q_1$ for all $q \in Q_P$, and $E(q) \subseteq Q_P$ for all $q \in Q_1$. The reduction is then as follows. Given an MDP $M = (Q, E, \delta)$ with a priority function $p : Q \to \mathbb{N}$ and a weight function $w : E \to \mathbb{Z}$, we construct $\langle M', p', w' \rangle$ as follows. M' is the MDP $M = (Q', E', \delta')$ where:

– $Q' = Q \cup (Q \times \{0, 2, \ldots, 2r\}) \cup \{\mathsf{sink}\}$ where $2r$ is the largest even priority of a state in Q. Intuitively, a state $(q, i) \in Q'$ corresponds to the state q of M from which player 1 will ensure to visit priority i (which is even) infinitely often, and never visit priority smaller than i;
– E' contains $E \cup \{(\mathsf{sink}, \mathsf{sink})\}$ and the following edges. For each probabilistic state $q \in Q_P$, for $i = 0, 2, \ldots, 2r$,
  - (a) if $p(q') \geq i$ for all $q' \in E(q)$, then $((q, i), (q', i)) \in E'$ for all $q' \in E(q)$,
  - (b) otherwise, $((q, i), \mathsf{sink}) \in E'$.

  For each player 1 state $q \in Q_1$, for each $q' \in E(q)$, for $i = 0, 2, \ldots, 2r$,
  - (a) $(q, \mathsf{sink}) \in E'$ and $((q, i), \mathsf{sink}) \in E'$, and
  - (b) if $p(q') \geq i$, then $(q, (q', i)) \in E'$ and $((q, i), (q', i)) \in E'$.

The partition (Q'_1, Q'_P) of Q' is defined by $Q'_1 = Q_1 \cup (Q_1 \times \{0, 2, \ldots, 2r\}) \cup \{\mathsf{sink}\}$ and $Q'_P = Q' \setminus Q'_1$. The weight of the edges (q, q'), $(q, (q', i))$ and $((q, i), (q', i))$ according to w' is the same as the weight of (q, q') according to w. The states (q, i) such that $p(q) = i$ have priority 0 according to p' (they are the Büchi states), and all the other states in Q' (including sink) have priority 1.

Lemma 3. *Given an MDP M with energy parity objective, we can construct in quadratic time an MDP M' with energy Büchi objective such that for all states q_0 in M, there exists an almost-sure winning strategy from q_0 in M if and only if there exists an almost-sure winning strategy from q_0 in M' (with the same initial credit).*

From the proof of Lemma 3, it follows that the memory requirement is the same as for energy Büchi MDPs. And if the weights are in $\{-1, 0, 1\}$, it follows that the energy parity problem can be solved in polynomial time.

Theorem 1. *For energy parity MDPs, (1) the decision problem of whether a given state is almost-sure winning is in NP $\cap$ coNP, and there is a pseudo-polynomial time algorithm in $O(|E| \cdot d \cdot |Q|^5 \cdot W)$ to solve it; and (2) memory of size $2 \cdot |Q| \cdot W$ is sufficient for almost-sure winning strategies.*

4 MDPs with Mean-payoff Parity Objectives

In this section we present a polynomial-time algorithm for solving MDPs with mean-payoff parity objective. We first recall some useful properties of MDPs.

For an end-component $U \in \mathcal{E}(M)$, consider the memoryless strategy σ_U that plays in every state $s \in U \cap Q_1$ all edges in $E(s) \cap U$ uniformly at random. Given the strategy σ_U, the end-component U is a closed connected recurrent set in the Markov chain obtained by fixing σ_U.

Lemma 4. *Given an MDP M and an end-component $U \in \mathcal{E}(M)$, the strategy σ_U ensures that for all states $s \in U$, we have $\mathbb{P}_s^{\sigma_U}(\{\omega \mid \mathsf{Inf}(\omega) = U\}) = 1$.*

Expected mean-payoff value. Given an MDP M with a weight function w, the *expected mean-payoff value*, denoted $\mathsf{ValMP}(w)$, is the function that assigns to every state the maximal expectation of the mean-payoff objective that can be guaranteed by any strategy. Formally, for $q \in Q$ we have $\mathsf{ValMP}(w)(q) = \sup_{\sigma \in \Sigma} \mathbb{E}_q^\sigma(\mathsf{MP}(w))$, where $\mathsf{MP}(w)$ is the measurable function that assigns to a play ρ the long-run average $\mathsf{MP}(w, \rho)$ of the weights. By the classical results of MDPs with mean-payoff objectives, it follows that there exists pure memoryless optimal strategies [15], i.e., there exists a pure memoryless optimal strategy σ^* such that for all $q \in Q$ we have $\mathsf{ValMP}(w)(q) = \mathbb{E}_q^{\sigma^*}(\mathsf{MP}(w))$.

It follows from Lemma 4 that the strategy σ_U ensures that from any starting state s, any other state t is reached in finite time with probability 1. Therefore, the value for mean-payoff parity objectives in MDPs can be obtained by computing values for end-components and then playing a strategy to maximize the expectation to reach the values of the end-components.

We now present the key lemma where we show that for an MDP that is an end-component such that the minimum priority is even, the mean-payoff parity objective $\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{\geq \nu}$ is satisfied with probability 1 if the expected mean-payoff value is at least ν at some state (the result also holds for strict inequality). In other words, from the expected mean-payoff value of at least ν we ensure that both the mean-payoff and parity objective is satisfied with probability 1 from all states. The proof of the lemma considers two pure memoryless strategies: one for stochastic shortest path and the other for optimal expected mean-payoff value, and combines them to obtain an almost-sure winning strategy for the mean-payoff parity objective (details in [8]).

Lemma 5. *Consider an MDP M with state space Q, a priority function p, and weight function w such that (a) M is an end-component (i.e., Q is an end-component) and (b) the smallest priority in Q is even. If there is a state $q \in Q$ such that $\mathsf{ValMP}(w) \geq \nu$ (resp. $\mathsf{ValMP}(w) > \nu$), then there exists a strategy σ^* such that for all states $q \in Q$ we have $\mathbb{P}_q^{\sigma^*}(\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{\geq \nu}) = 1$ (resp. $\mathbb{P}_q^{\sigma^*}(\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{> \nu}) = 1$).*

Memory required by strategies. Lemma 5 shows that if the smallest priority in an end-component is even, then considering the sub-game restricted to the end-component, the mean-payoff parity objective is satisfied if and only if the mean-payoff objective is satisfied. The strategy constructed in Lemma 5 requires infinite memory, and in the case of loose inequality (i.e., $\mathsf{MeanPayoff}^{\geq \nu}$) infinite memory is required in general (see [11] for an example on graphs), and if the inequality is strict (i.e., $\mathsf{MeanPayoff}^{> \nu}$), then finite memory strategies exist [17]. For the purpose of computation we show that both strict and non-strict inequality can be solved in polynomial time. Since Lemma 5 holds for both strict and non-strict inequality, in sequel of this section we consider non-strict inequality and all the results hold for strict inequality as well.

Winning end-component. Given an MDP M with a parity objective $\mathsf{Parity}(p)$ and a mean-payoff objective $\mathsf{MeanPayoff}^{\geq \nu}$ for a weight function w, we call an end-component U *winning* if (a) $\min(p(U))$ is even; and (b) there exists a state with expected mean-payoff value at least ν in the sub-MDP induced by U, i.e., $\max_{q \in U} \mathsf{ValMP}(w)(q) \geq \nu$ in the sub-MDP induced by U. We denote by $\mathcal{W}$ the set of winning end-components, and let $\mathsf{Win} = \bigcup_{U \in \mathcal{W}} U$ be the union of the winning end-components.

Reduction to reachability of winning end-component. By Lemma 5 it follows that in every winning end-component the mean-payoff parity objective is satisfied with probability 1. Conversely, consider an end-component U that is not winning, then either the smallest priority is odd, or the maximal expected mean-payoff value that can be ensured for any state in U by staying in U is less than ν. Hence if only states in U are visited infinitely often, then with probability 1 (i) either the parity objective is not satisfied, or (ii) the mean-payoff objective is not satisfied. In other words, if an end-component that is not winning is visited infinitely often, then the mean-payoff parity objective is satisfied with probability 0. It follows that the value function for MDPs with mean-payoff parity objective can be computed by computing the value function for reachability to the set Win, i.e., formally, $\sup_{\sigma \in \Sigma} \mathbb{P}_q^\sigma(\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{\geq \nu}) = \sup_{\sigma \in \Sigma} \mathbb{P}_q^\sigma(\mathsf{Reach}(\mathsf{Win}))$, where $\mathsf{Reach}(\mathsf{Win})$ is the set of paths that reaches a state in Win at least once. Since the value function in MDPs with reachability objectives can be computed in polynomial time using linear programming [15], it suffices to present a polynomial-time algorithm to compute Win in order to obtain a polynomial-time algorithm for MDPs with mean-payoff parity objectives.

Computing winning end-components. The computation of the winning end-components is done iteratively by computing winning end-components with smallest priority 0, then winning end-components with smallest priority 2, and so on. The computation of Win is as follows:

- For $i \geq 0$, let $\mathcal{W}_{2i}$ be the set of maximal end-components U with states with priority at least $2i$ and that contain at least one state with priority $2i$, i.e., U contains only states with priority at least $2i$, and contains at least one state with priority $2i$. Let $\mathcal{W}'_{2i} \subseteq \mathcal{W}_{2i}$ be the set of maximal end-components $U \in \mathcal{W}_{2i}$ such that there is a state $q \in U$ such that the expected mean-payoff value in the sub-MDP restricted to U is at least ν. Let $\mathsf{Win}_{2i} = \bigcup_{U \in \mathcal{W}'_{2i}} U$.

The set $\mathsf{Win} = \bigcup_{i=0}^{\lfloor d/2 \rfloor} \mathsf{Win}_{2i}$ is the union of the states of the winning end-components (formal pseudo-code in [8]).

Complexity of computing winning end-components. The winning end-component algorithm runs for $O(d)$ iterations and in each iteration requires to compute a maximal end-component decomposition and compute mean-payoff values of at most n end-components, where n is the number of states of the MDP. The maximal end-component decomposition can be achieved in polynomial time [12, 13, 9]. The mean-payoff value function of an MDP can also be computed in polynomial time using linear programming [15]. It follows that the value function of an MDP with mean-payoff parity objectives can be computed in polynomial time. The almost-sure winning set is obtained by computing almost-sure reachability to Win in polynomial time [12, 13, 9]. This polynomial-time complexity provides a tight upper bound for the problem.

Theorem 2. *The following assertions hold:*

1. *The set of almost-sure winning states for mean-payoff parity objectives can be computed in polynomial time for MDPs.*
2. *For mean-payoff parity objectives, almost-sure winning strategies require infinite memory in general for non-strict inequality (i.e, for mean-payoff parity objectives* $\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{\geq \nu}$*) and finite-memory almost-sure winning strategies exist for strict inequality (i.e., for* $\mathsf{Parity}(p) \cap \mathsf{MeanPayoff}^{> \nu}$*).*

References

1. A. Bianco and L. de Alfaro. Model checking of probabilistic and nondeterministic systems. In *FSTTCS 95*, volume 1026 of *LNCS*, pages 499–513, 1995.
2. R. Bloem, K. Chatterjee, T. A. Henzinger, and B. Jobstmann. Better quality in synthesis through quantitative objectives. In *CAV*, LNCS 5643, pages 140–156. Springer, 2009.
3. P. Bouyer, U. Fahrenberg, K. G. Larsen, N. Markey, and J. Srba. Infinite runs in weighted timed automata with energy constraints. In *Proc. of FORMATS*, LNCS 5215, pages 33–47. Springer, 2008.
4. Tomás Brázdil, Václav Brozek, Kousha Etessami, Antonín Kucera, and Dominik Wojtczak. One-counter Markov decision processes. In *Proc. of SODA*, pages 863–874. SIAM, 2010.
5. L. Brim, J. Chaloupka, L. Doyen, R. Gentilini, and J.-F. Raskin. Faster algorithms for mean-payoff games. *Formal Methods in System Design*, 2010.
6. A. Chakrabarti, L. de Alfaro, T. A. Henzinger, and M. Stoelinga. Resource interfaces. In *Proc. of EMSOFT*, LNCS 2855, pages 117–133. Springer, 2003.
7. K. Chatterjee and L. Doyen. Energy parity games. In *Proc. of ICALP: Automata, Languages and Programming (Part II)*, LNCS 6199, pages 599–610. Springer, 2010.
8. K. Chatterjee and L. Doyen. Energy and mean-payoff parity Markov decision processes. Technical report, IST Austria, Feb, 2011. `http://pub.ist.ac.at/Pubs/TechRpts/2011/IST-2011-0001.pdf`.
9. K. Chatterjee and M. Henzinger. Faster and dynamic algorithms for maximal end-component decomposition and related graph problems in probabilistic verification. In *Proc. of SODA*. ACM SIAM, 2011.
10. K. Chatterjee, T. A. Henzinger, B. Jobstmann, and R. Singh. Measuring and synthesizing systems in probabilistic environments. In *CAV*, pages 380–395, 2010.
11. K. Chatterjee, T. A. Henzinger, and M. Jurdziński. Mean-payoff parity games. In *Proc. of LICS*, pages 178–187. IEEE Computer Society, 2005.
12. C. Courcoubetis and M. Yannakakis. The complexity of probabilistic verification. *J. ACM*, 42(4):857–907, 1995.
13. L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, 1997.
14. E. A. Emerson and C. Jutla. Tree automata, mu-calculus and determinacy. In *Proc. of FOCS*, pages 368–377. IEEE, 1991.
15. J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1997.
16. H. Gimbert and F. Horn. Solving simple stochastic tail games. In *Proc. of SODA*, pages 847–862, 2010.
17. H. Gimbert, Y. Oualhadj, and S. Paul. Computing optimal strategies for Markov decision processes with parity and positive-average conditions. Technical report, LaBRI, Université de Bordeaux II, 2011.
18. M. Jurdziński. Deciding the winner in parity games is in UP $\cap$ co-UP. *Inf. Process. Lett.*, 68(3):119–124, 1998.
19. A. Kucera and O. Stražovský. On the controller synthesis for finite-state markov decision processes. In *Proc. of FSTTCS*, pages 541–552, 2005.
20. A. Pacuk. Hybrid of mean payoff and total payoff. In *Talk at the workshop "Games for Design and Verification, St Anne's College Oxford"*, September 2010.
21. W. Thomas. Languages, automata, and logic. In *Handbook of Formal Languages*, volume 3, Beyond Words, chapter 7, pages 389–455. Springer, 1997.
22. M.Y. Vardi. Automatic verification of probabilistic concurrent finite-state systems. In *FOCS'85*. IEEE Computer Society Press, 1985.
23. U. Zwick and M. Paterson. The complexity of mean payoff games on graphs. *Theor. Comput. Sci.*, 158(1&2):343–359, 1996.