# Improving Synthesis of Finite State Controllers for POMDPs Using Belief Space Approximation

**Ing. Filip Macák** (imacak@fit.vutbr.cz)

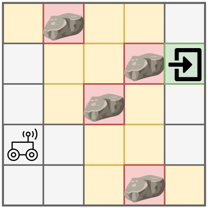
Supervisor: Assoc. Prof. Milan Češka

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

# Problem Formulation

**Partially-observable Markov decision processes** (POMDPs)

- prominent model for sequential decision-making under uncertainty and limited observability
- observations – states with the same observation are indistinguishable
- observation-based policies are required



observations:
- exit
- crashed
- near obstacle
- other

**Specification**:

- minimise the number of steps to reach the exit
- keep the probability of crashing below 1%

**Many practical applications:**

- planning of autonomous agents and robotics
- games with imperfect information (e.g texas holdem)
- medical treatment strategies (e.g heart disease)

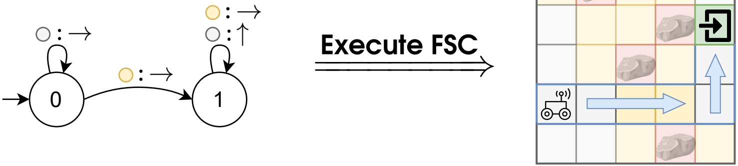
# Problem Formulation

Find the optimal policy for the given infinite-horizon specifications

- no discounting – much harder than finite-horizon problems
- important for long-term planning and sparse-rewards problems
- in general undecidable – policy may require infinite memory

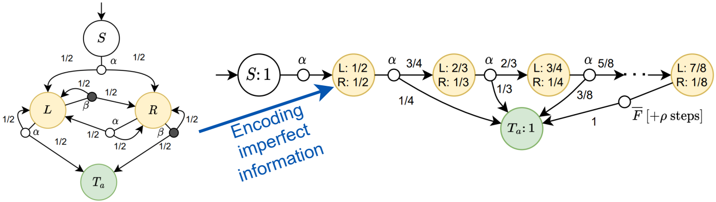
We aim at compact, verifiable and easy-to-execute strategies

- **finite-state controller** (FSC) based on Mealy machines




Anytime algorithm: in the given time, find the best FSC

Belief - probability distribution over the states of a POMDP




Construct and analyse the reachable belief space

- it might be huge or even infinite
- various approximations of the unexplored belief space, namely, cut-offs[1] and point-based

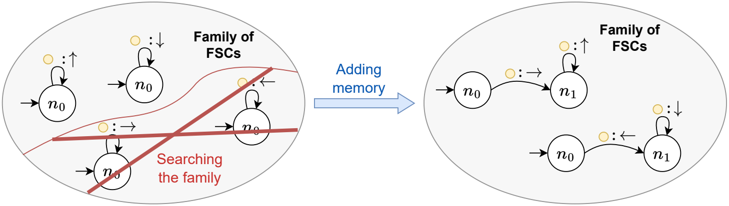
Limitations:

- cut-offs (implemented in the tool Storm) are not sufficient
- point-based methods, notably SARSOP[2], perform poorly for long-term planning

[1] A. Bork et al. Under-approximating expected total rewards in POMDPs. In TACAS'22.
[2] H. Kurniawati et al. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In Robotics: Science and Systems 2008.

Symbolic representation and exploration of families of candidate FSCs

- fully-observable abstraction and counter-examples steer the exploration
- iterative expansion of the family by adding memory nodes
- implemented in the tool PAYNT[3] (developed at BUT FIT)



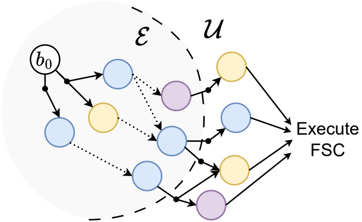

Limitations:

- the family size grows exponentially with the memory
- if a lot of memory is needed or the POMDP is too large, exploration becomes computationally intractable

[3] R. Andriushchenko et al. Inductive synthesis of finite-state controllers for POMDPs. In UAI'22.

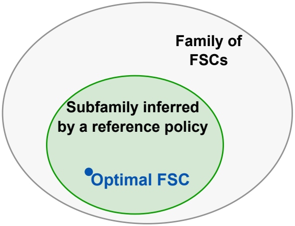
## Two novel ideas

Using FSCs as cut-offs to obtain a better approximation of the unexplored parts of the belief space




Already very non-optimal FSCs improve bounds provided by existing cut-offs techniques

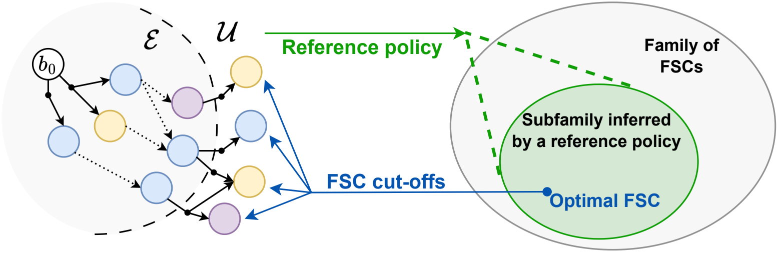
Using reference policies from belief-space exploration to guide the inductive synthesis




Already very shallow exploration of the belief space is useful for guiding family exploration

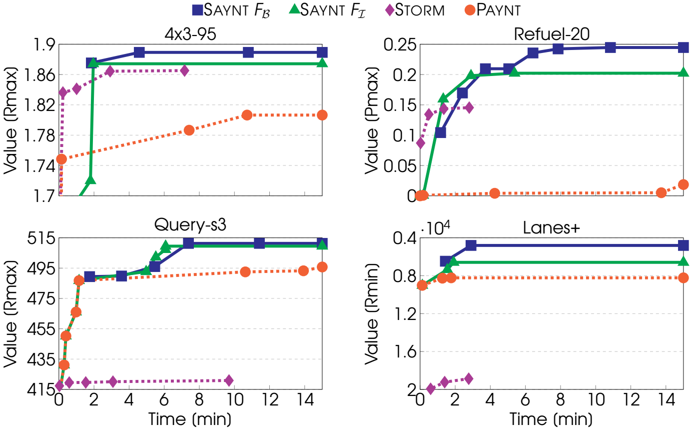
**SAYNT** is an iterative anytime synthesis algorithm

- closed-loop integration of the inductive synthesis and the belief-space exploration
  - PAYNT provides cut-off FSCs for Storm,
  - Storm provides reference policies for PAYNT and suggest where to add the memory
- in each iteration two FSCs $F_{\mathcal{I}}$ and $F_{\mathcal{B}}$ are obtained



# Experimental Evaluation

## Comparing SAYNT and state-of-the-art tools Storm[1] and PAYNT[3]

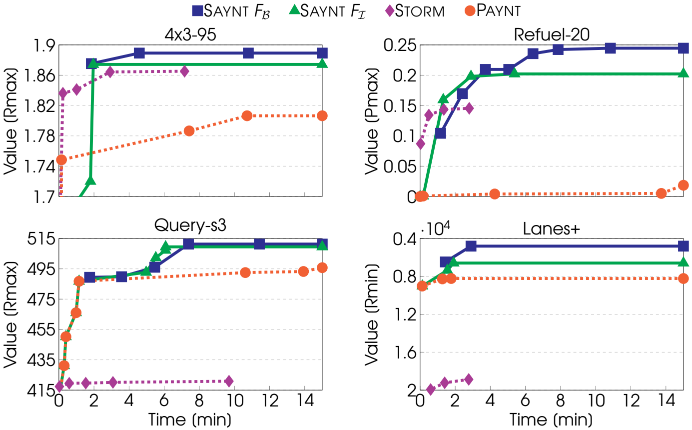



**SAYNT steadily outperforms both baselines on a wide range of benchmarks from AI and formal verification communities**

[1] A. Bork et al. Under-approximating expected total rewards in POMDPs. In TACAS'22.

[3] R. Andriushchenko et al. Inductive synthesis of finite-state controllers for POMDPs. In UAI'22.

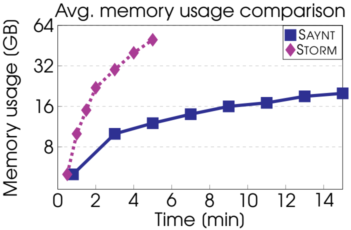
Comparing SAYNT and state-of-the-art tools Storm[1] and PAYNT[3]




**The quality of improvements grows with the complexity of POMDPs and reaches up to 40%**

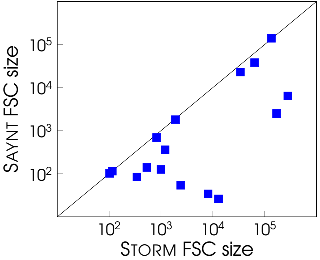
[1] A. Bork et al. Under-approximating expected total rewards in POMDPs. In TACAS'22.
[3] R. Andriushchenko et al. Inductive synthesis of finite-state controllers for POMDPs. In UAI'22.

## Memory footprint




- SAYNT significantly reduces memory usage compared to Storm
- This allows an efficient belief-space exploration in larger POMDPs

## FSC size comparison




- SAYNT produces more compact FSCs compared to Storm while achieving better values

**Conference paper based on this work has been accepted to CAV'23 (A* conference)**

I would like to thank all the co-authors:

- Milan Češka (**BUT FIT**)
- Roman Andriushchenko (**BUT FIT**)
- Alexander Bork (RWTH Aachen University)
- Sebastian Junges (Radboud University)
- Joost-Pieter Katoen (RWTH Aachen University)

My key contributions:

- Formulation of research ideas, namely improvements to the inductive synthesis and the idea of a symbiotic approach
- Design and implementation of the enhanced inductive synthesis and of the symbiotic loop
- Implementation of the export of belief policies
- Experimental evaluation and artifact preparation
- Writing of the paper

# Conclusions

**Novel algorithm for POMDPs and infinite-horizon specifications**

- symbiotically integrates the belief-space exploration and the inductive synthesis
- outperforms state-of-the-art methods on a wide range of benchmarks

**Future research:**

- discounted vs. undiscounted specifications
- extension to partially-observable stochastic games
- combination with RL-based approaches