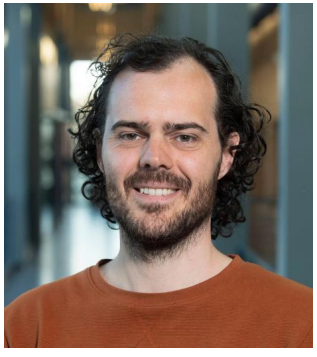
Shielded Reinforcement Learning for Hybrid Systems

Christian Schilling


joint work with Asger Horn Brorholt, Peter Gjøl Jensen,
Kim Guldstrand Larsen, and Florian Lorber

July 20, 2023

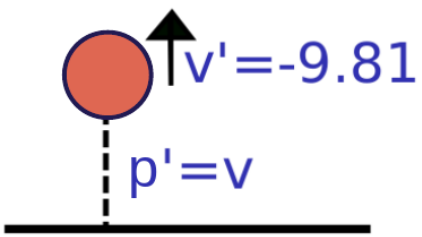
AALBORG UNIVERSITET

Overview

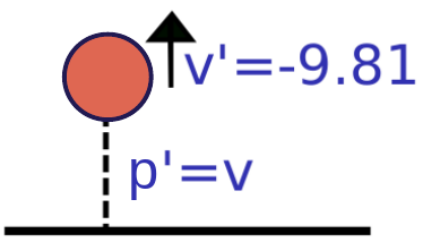
Overview

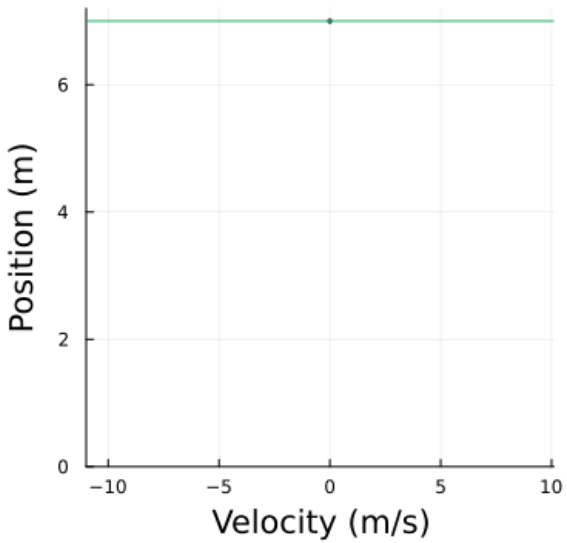
Control of physical systems

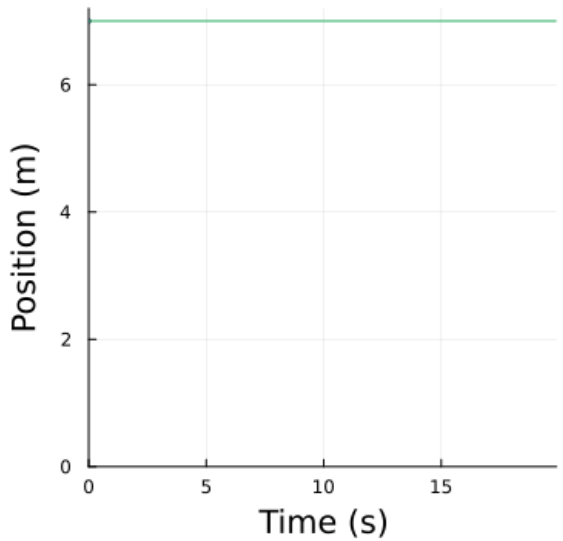
- Physical systems have complex dynamics
  - Continuous evolution
  - Discrete events
  - Stochastic uncertainty



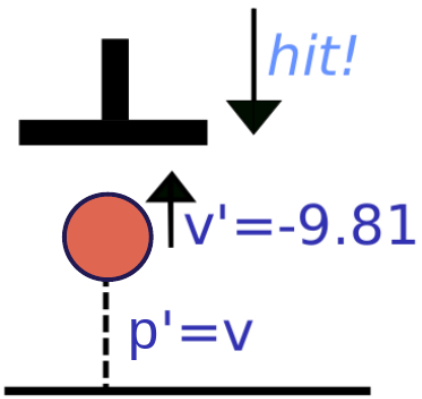
Control of physical systems





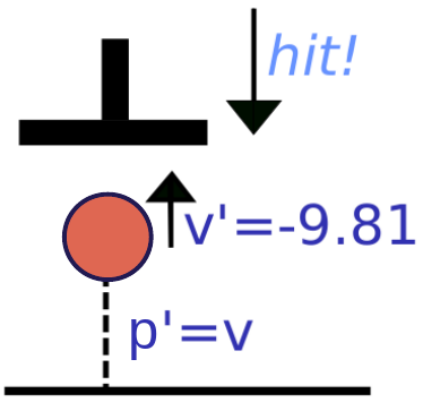


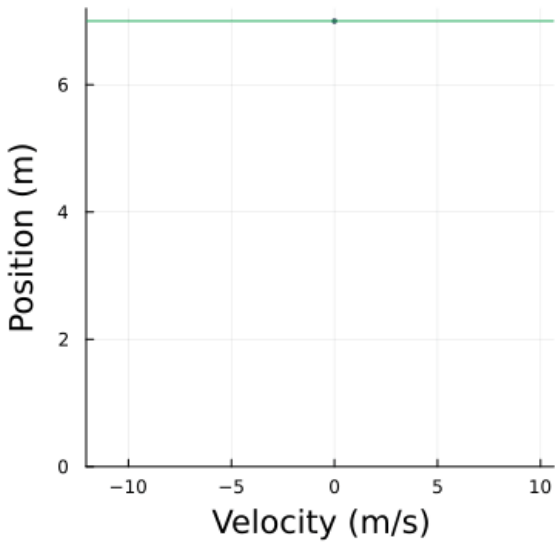
Control of physical systems

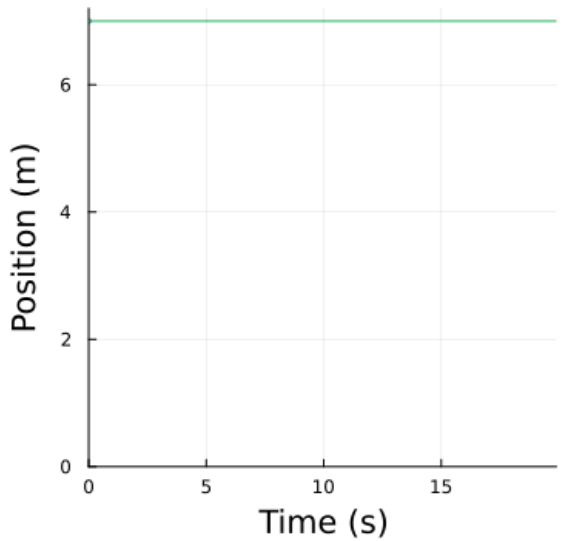


- Physical systems have complex dynamics
  - Continuous evolution
  - Discrete events
  - Stochastic uncertainty
- **We want to control these systems subject to some optimality criterion, e.g., minimize number of hits**

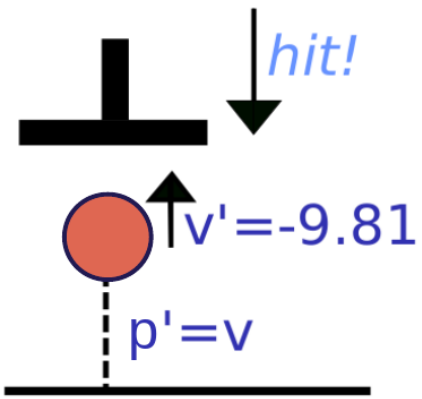
Control of physical systems





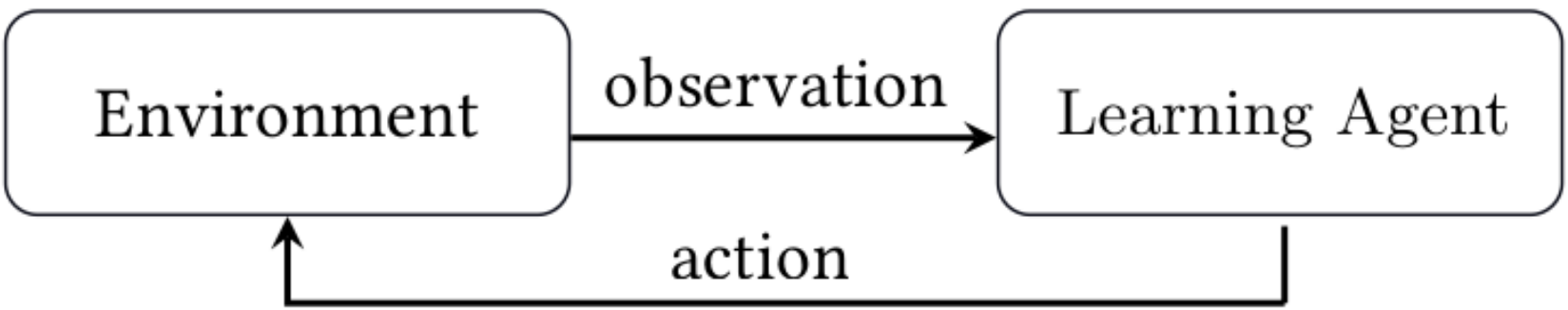


Control of physical systems

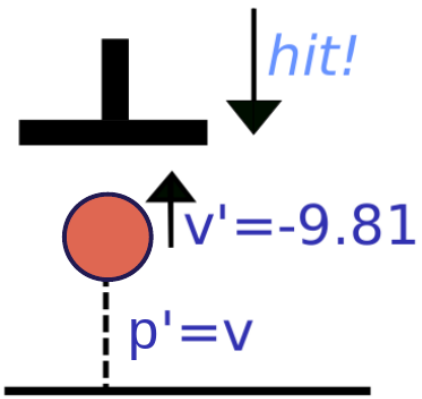


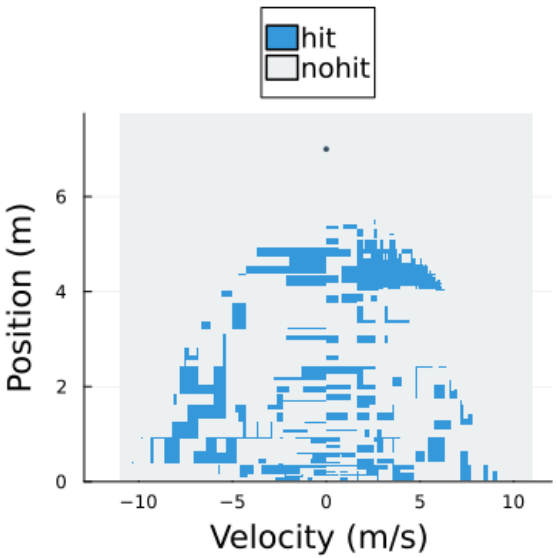
- Physical systems have complex dynamics
  - Continuous evolution
  - Discrete events
  - Stochastic uncertainty
- We want to control these systems subject to some optimality criterion, e.g., minimize number of hits
- **We can reinforcement-learn a controller, e.g., with Uppaal Stratego**

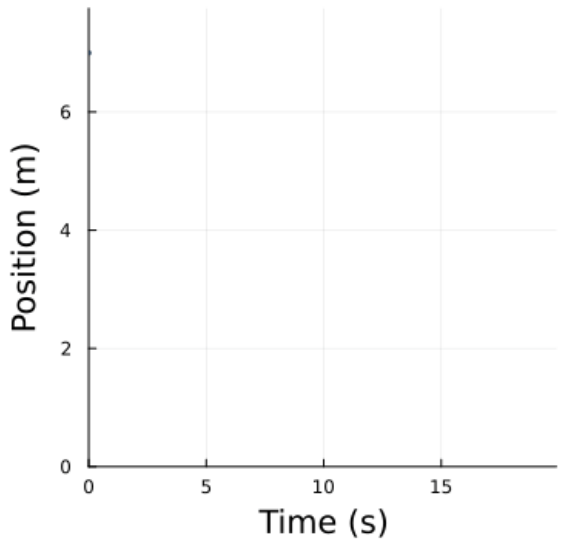
Reinforcement learning



Control of physical systems

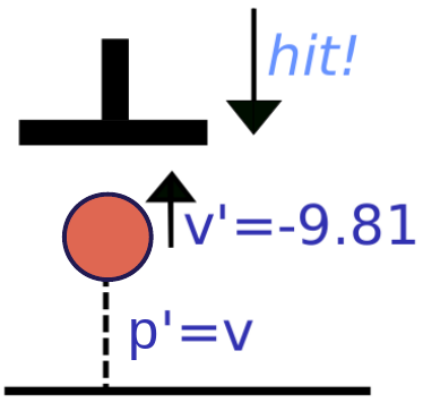






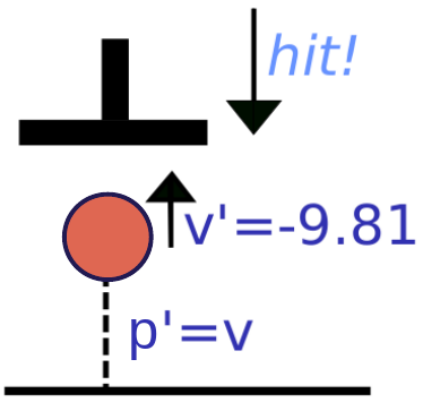
Trained for 12,000 episodes

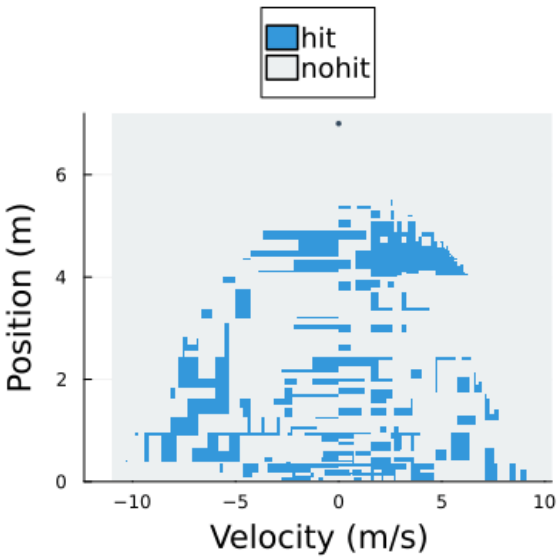
Control of physical systems

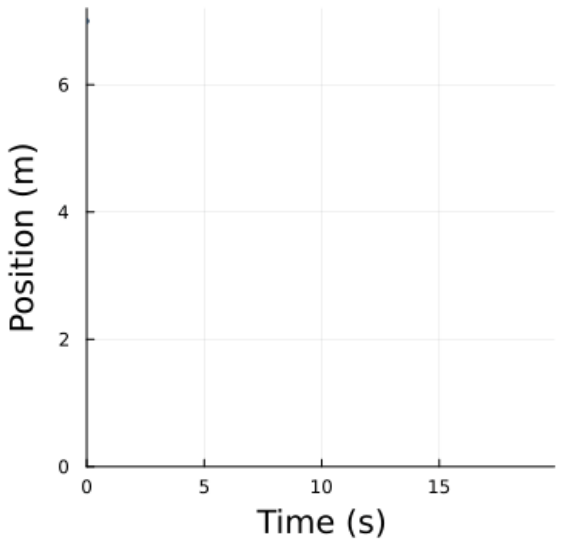


- Physical systems have complex dynamics
  - Continuous evolution
  - Discrete events
  - Stochastic uncertainty
- We want to control these systems subject to some optimality criterion,
  e.g., minimize number of hits
- We can reinforcement-learn a controller,
  e.g., with Uppaal Stratego
- **We also have safety constraints,**
  e.g., $p = 0 \implies |v| > 1$

Control of physical systems



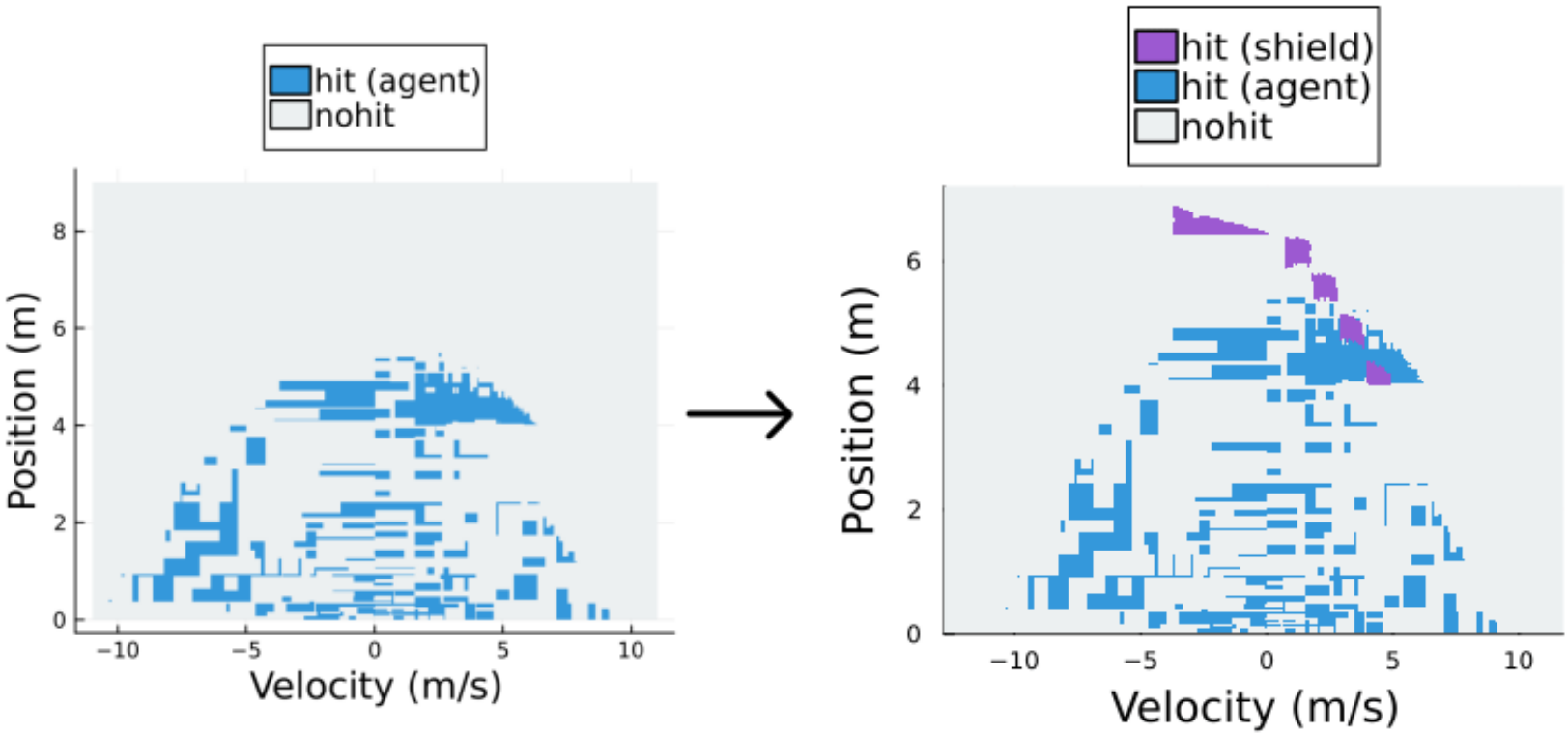




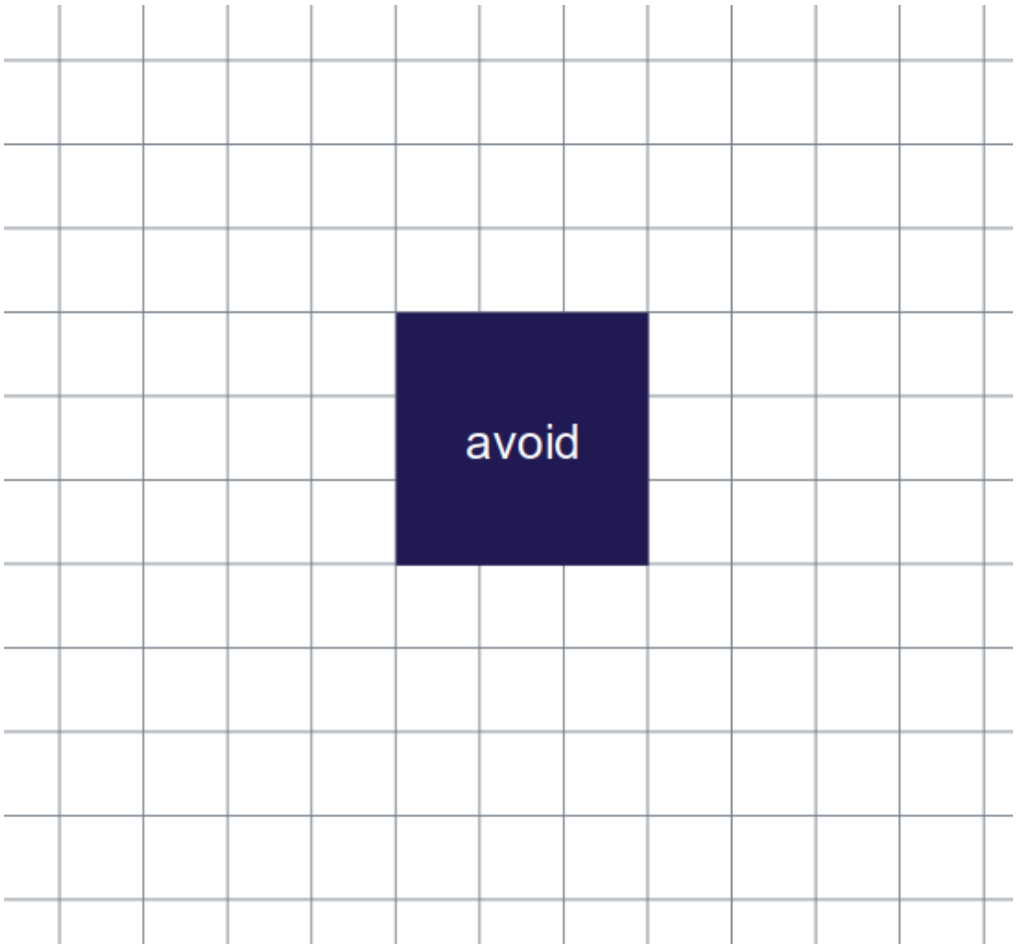
2 % of executions unsafe

Overview

Shielding

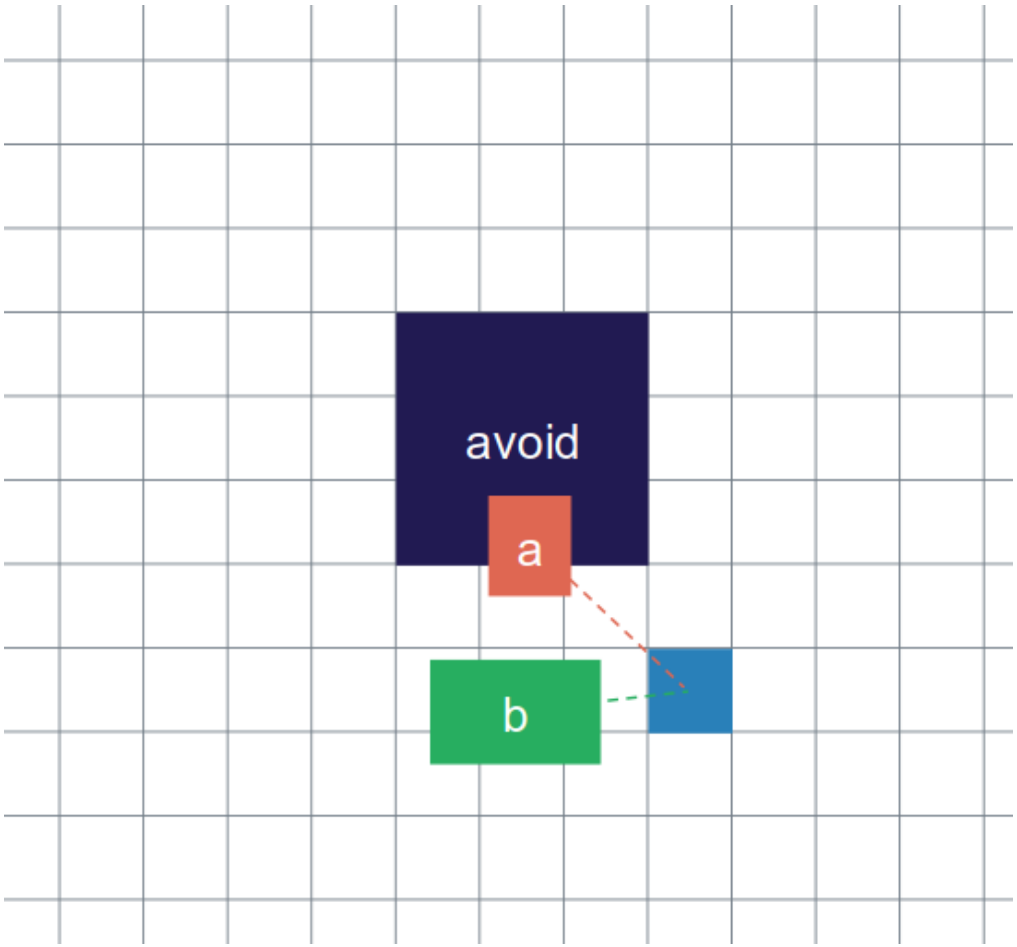

hit (agent)
nohit
Position (m)
Velocity (m/s)
hit (shield)
hit (agent)
nohit
Position (m)
Velocity (m/s)

State-space partitioning and two-player game



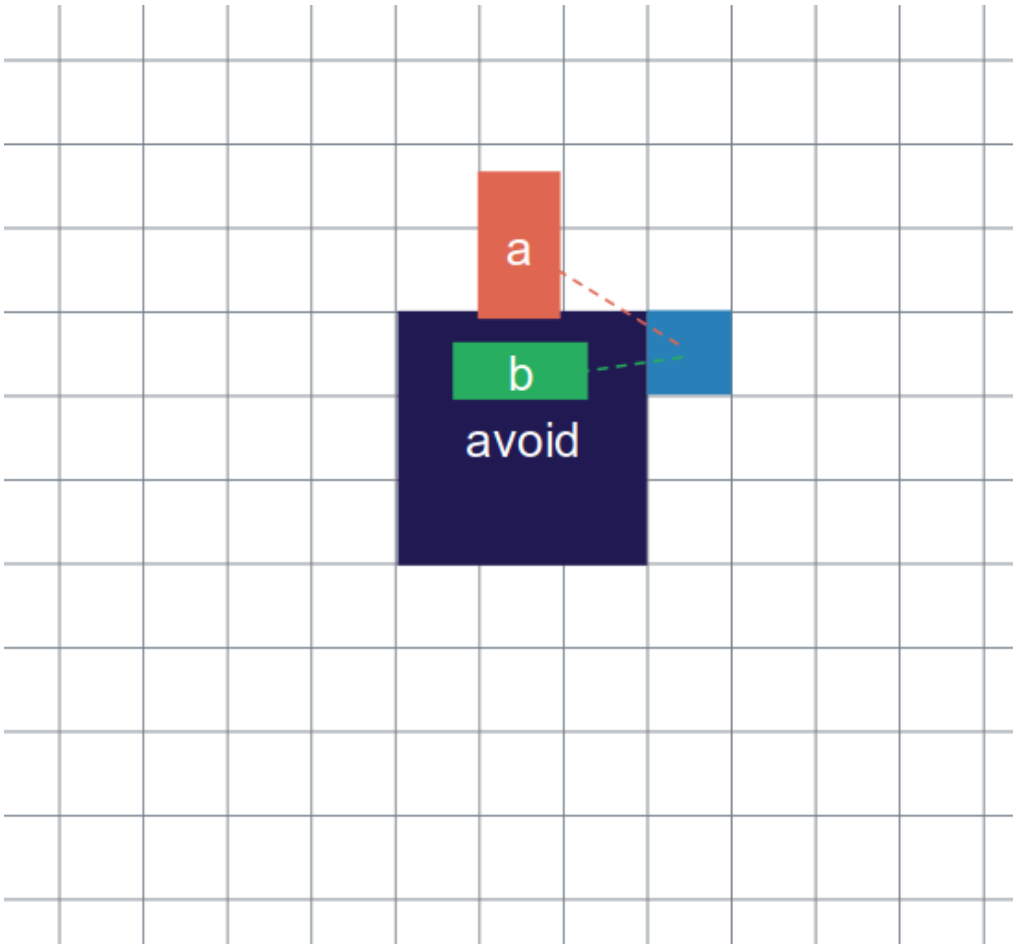
- For each block, compute the successor states under each action

State-space partitioning and two-player game



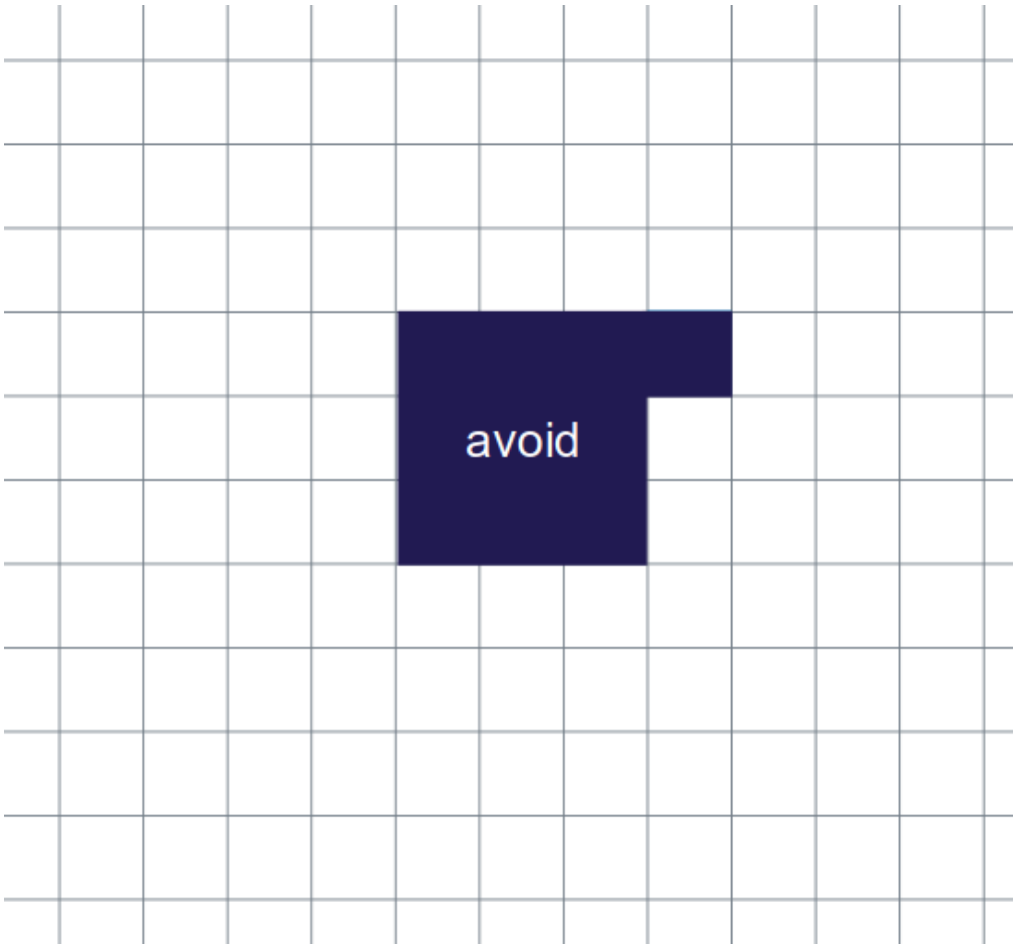
- For each block, compute the successor states under each action
- Eliminate actions that reach the avoid set

State-space partitioning and two-player game



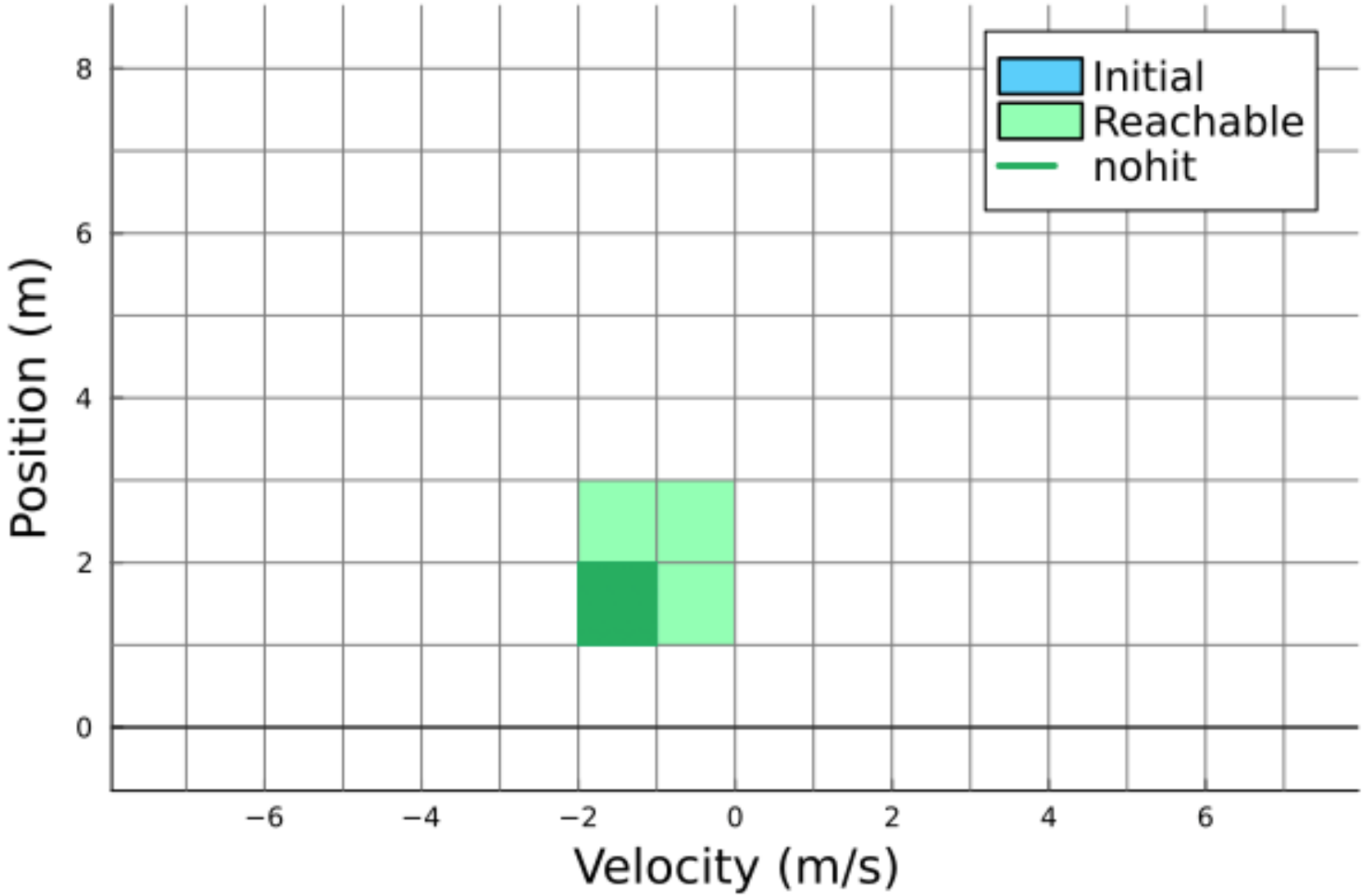
- For each block, compute the successor states under each action
- Eliminate actions that reach the avoid set
- If no action is safe,

State-space partitioning and two-player game

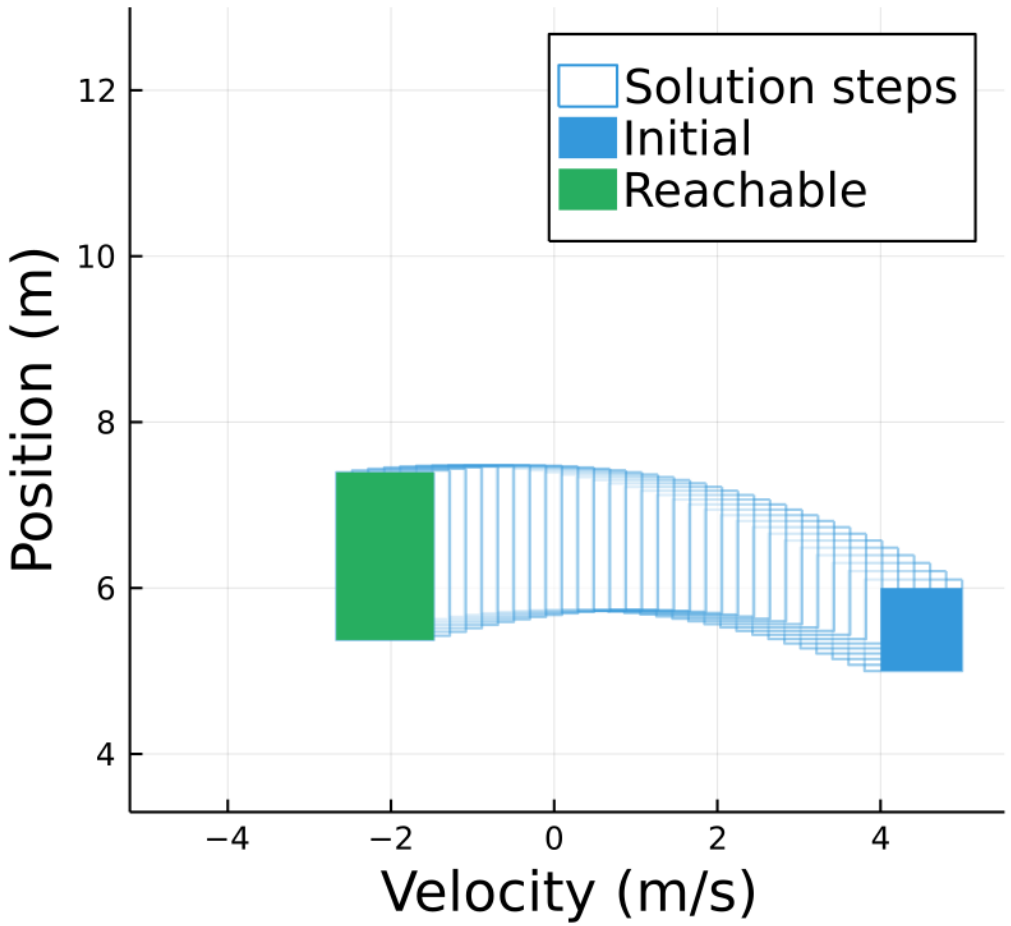


- For each block, compute the successor states under each action
- Eliminate actions that reach the avoid set
- If no action is safe, add the block to the avoid set

How to compute reachable blocks for complex systems?

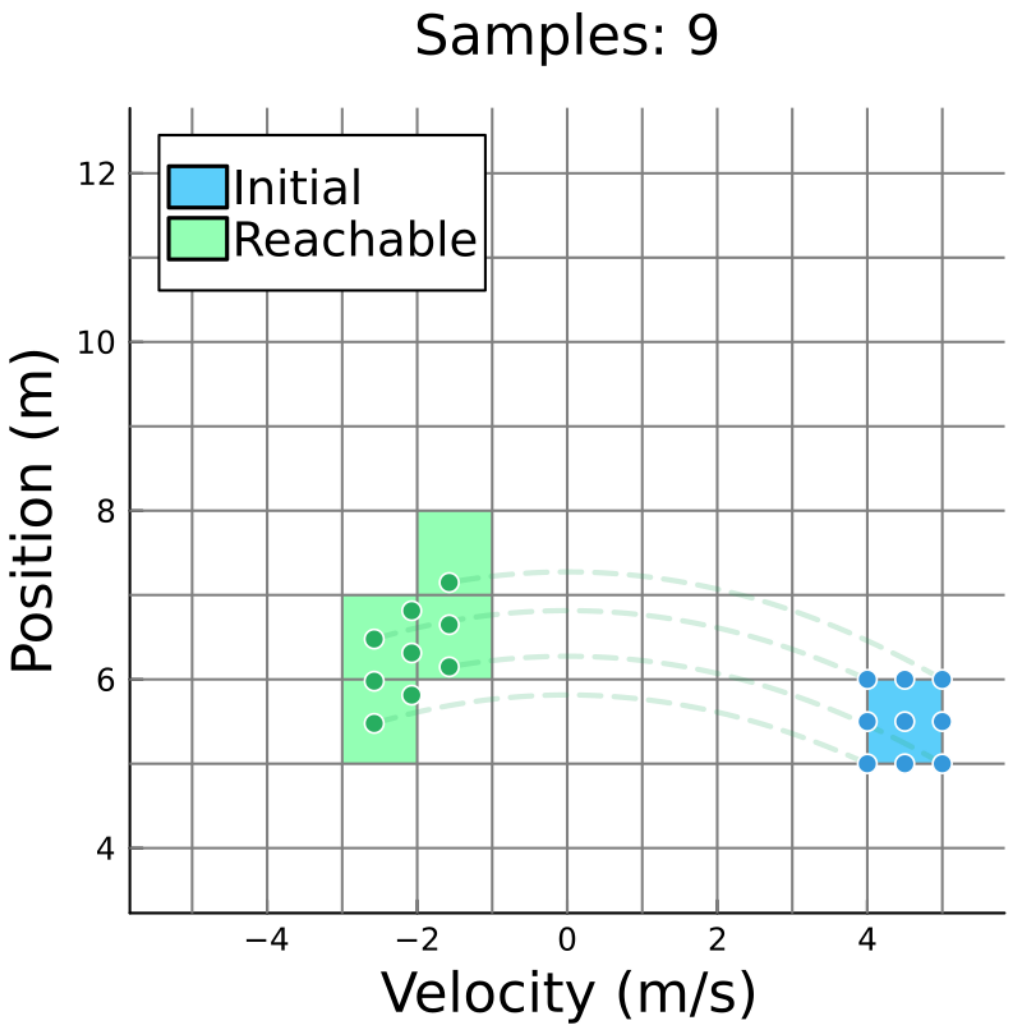


Computation via reachability method (here: JuliaReach)



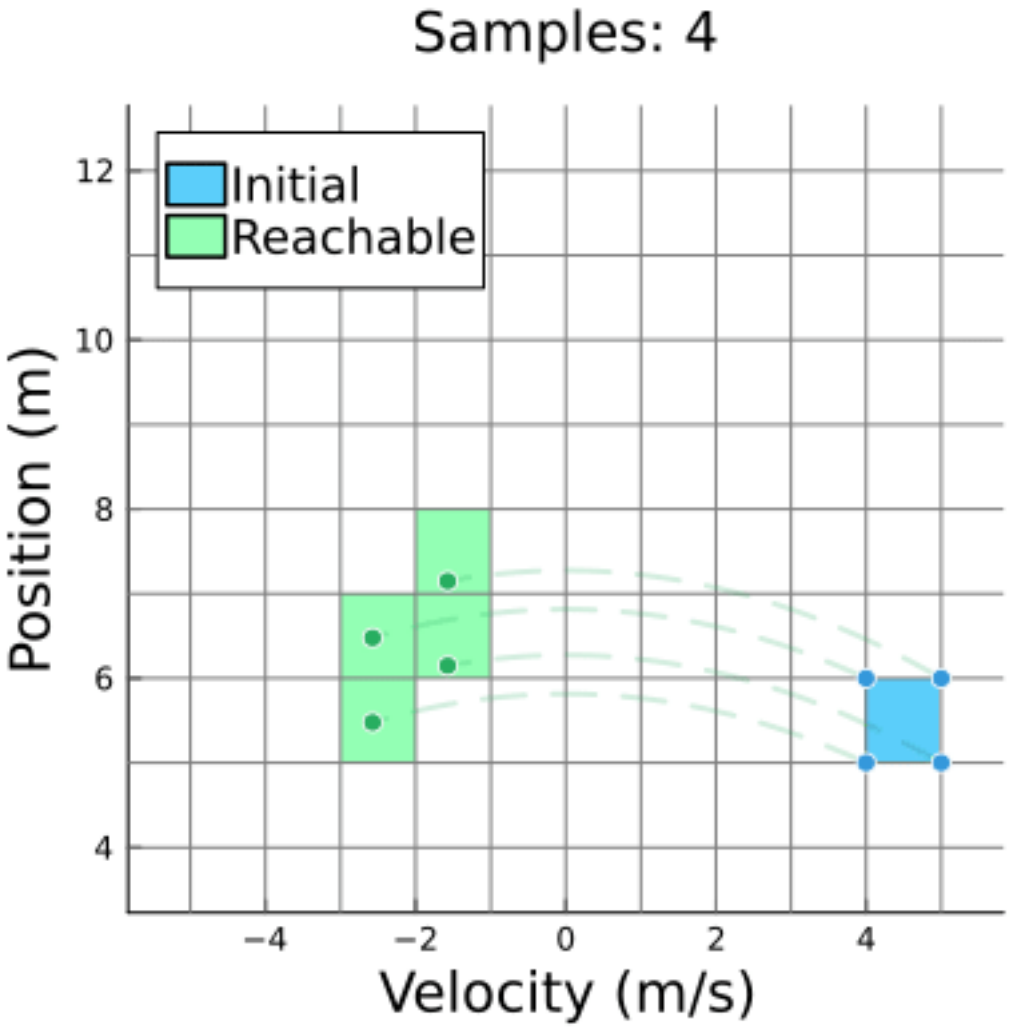
- Sound result
- Expensive

Computation via simulation-based method



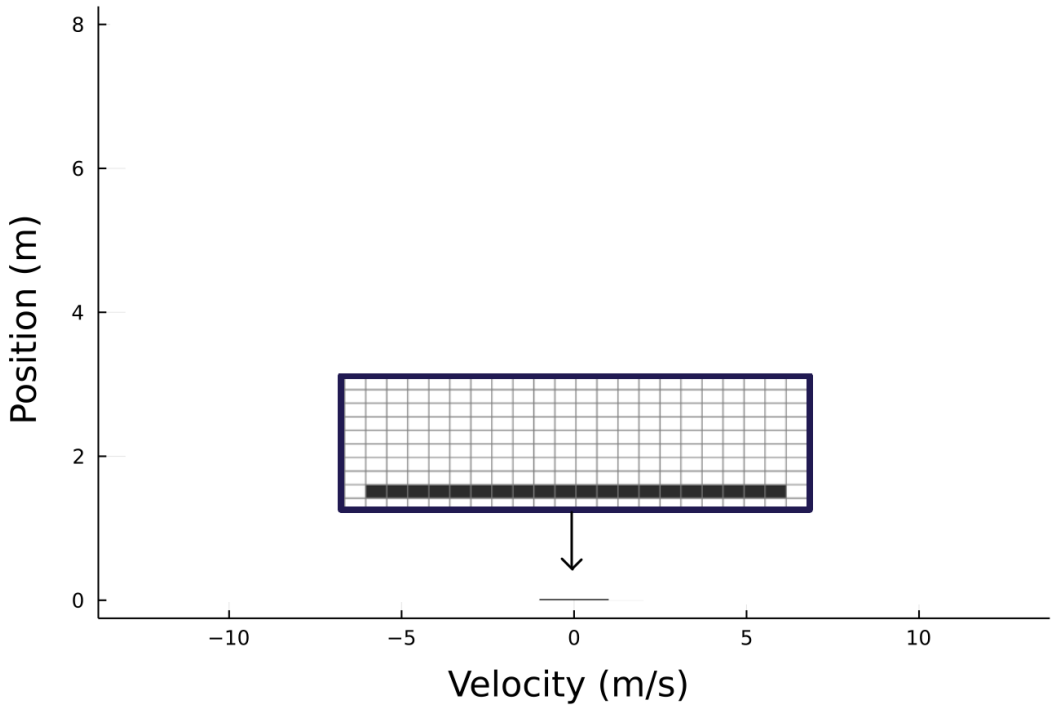
- Cheap
- May miss behavior
- Accuracy improves with more samples

Computation via simulation-based method



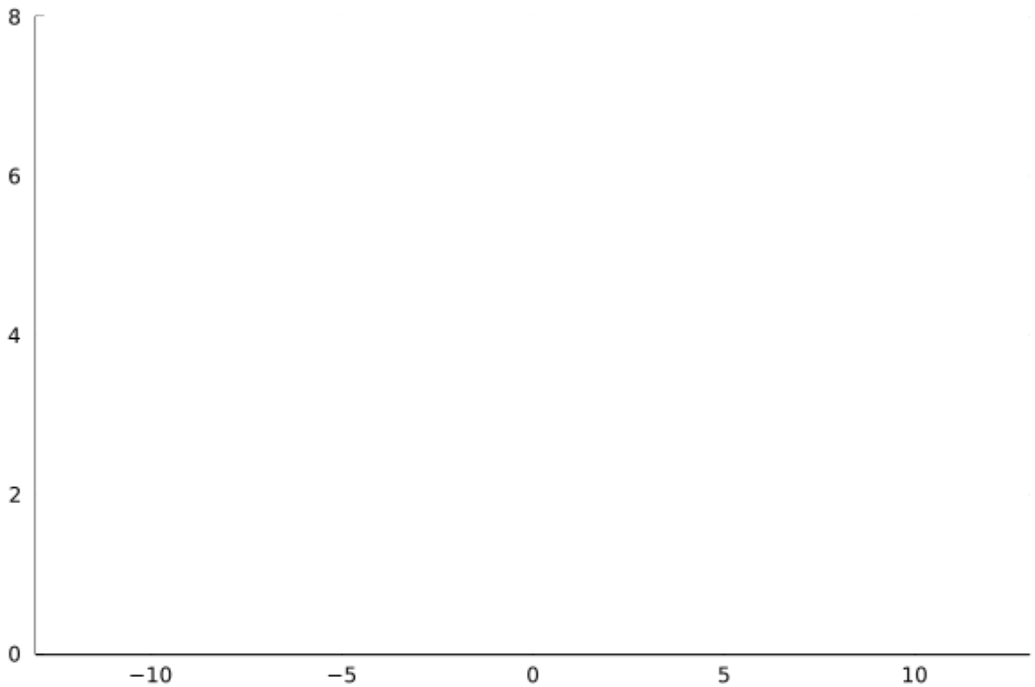
- Cheap
- May miss behavior
- Accuracy improves with more samples

Synthesis algorithm in action



- 16 samples per block
- Grid size 0.02
- Synthesis: 134 sec

Synthesis algorithm in action



- 16 samples per block
- Grid size 0.02
- Synthesis: 134 sec

Overview

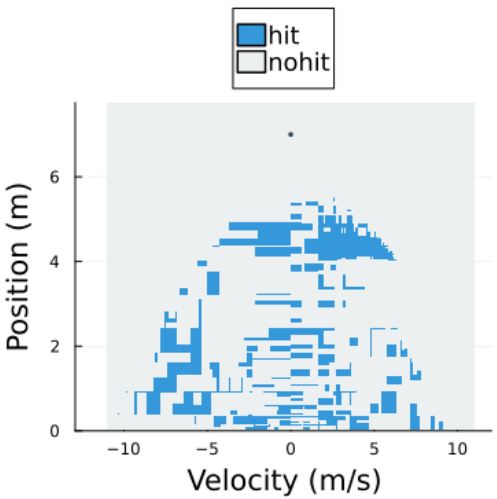
Scalability

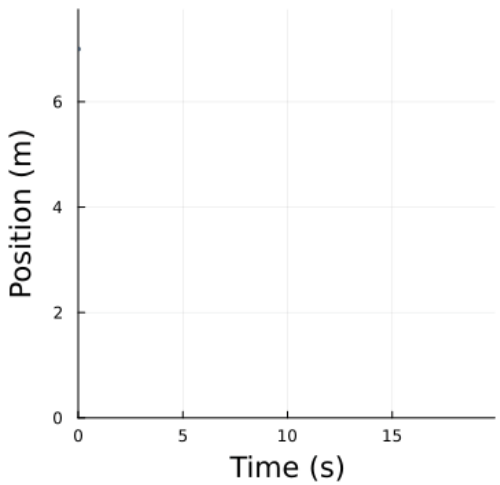
Grid size	Samples	Time
0.02	4	2m 14s
0.02	8	4m 02s
0.02	16	11m 03s
0.01	4	19m 00s
0.01	8	27m 21s
0.01	16	56m 32s

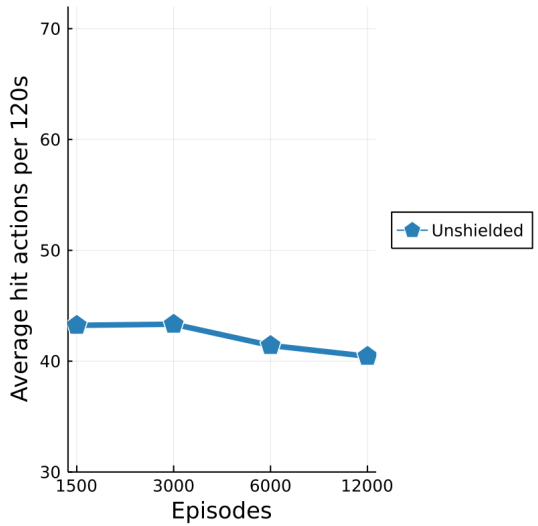
Grid size	Time
0.01	41h 05m

- Statistically safe ($\geq 99.99\,\%$ with confidence $99\,\%$)

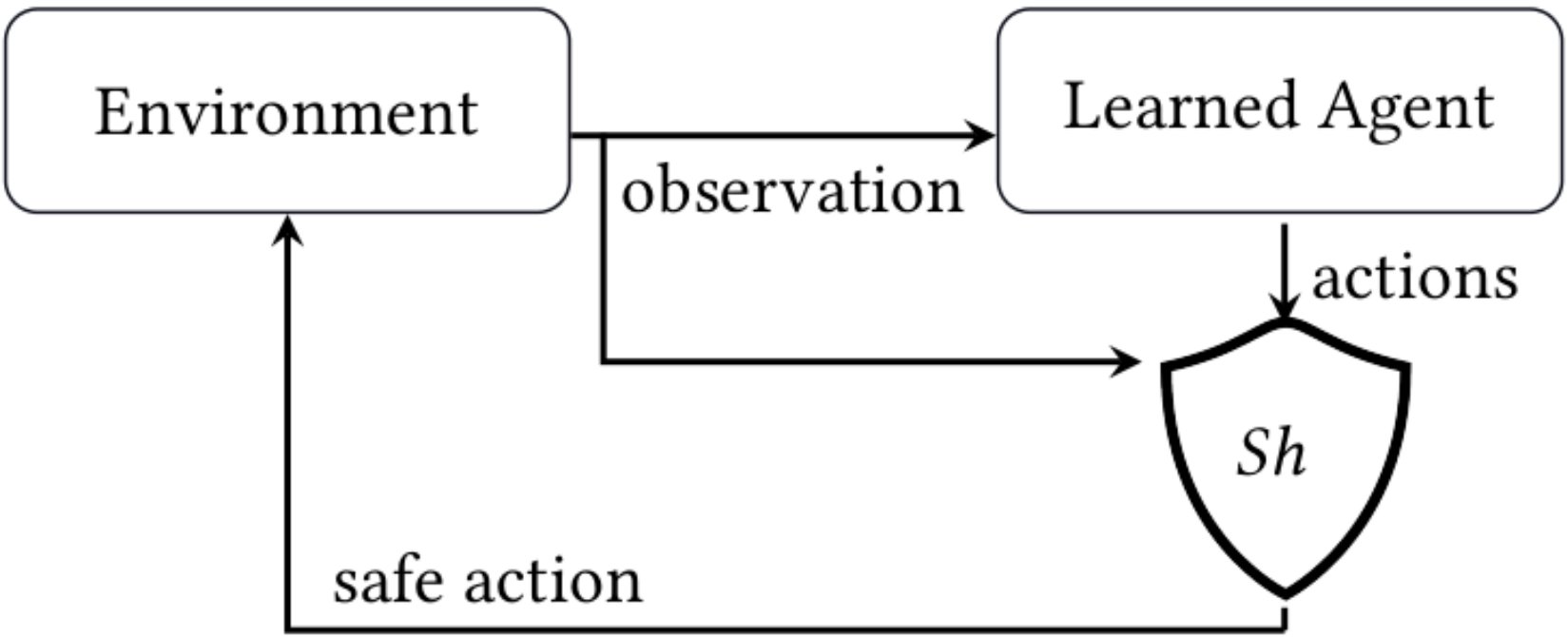
Unshielded agent



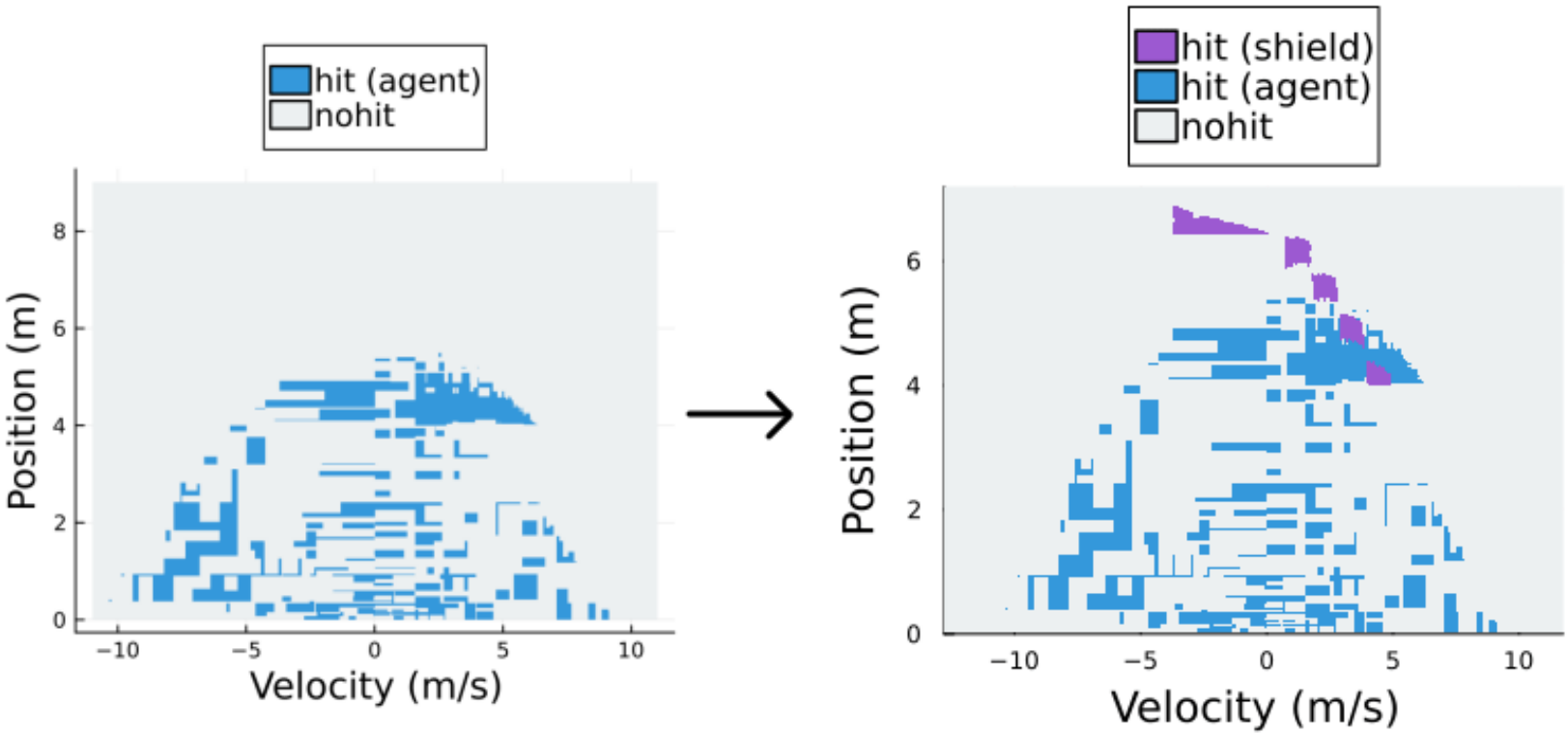




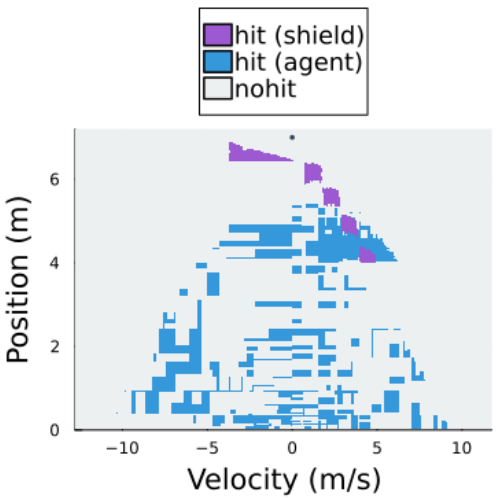
Post-shielded agent

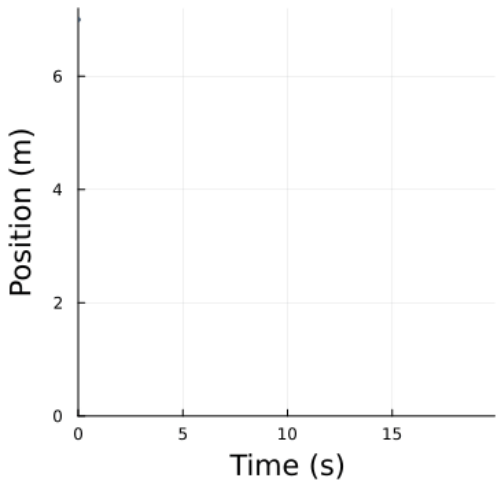


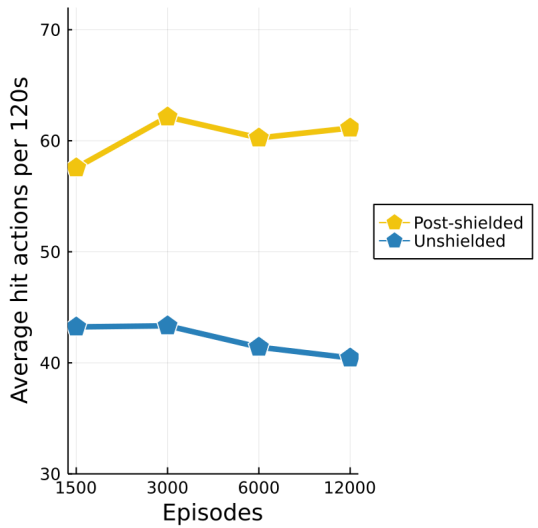
Post-shielded agent



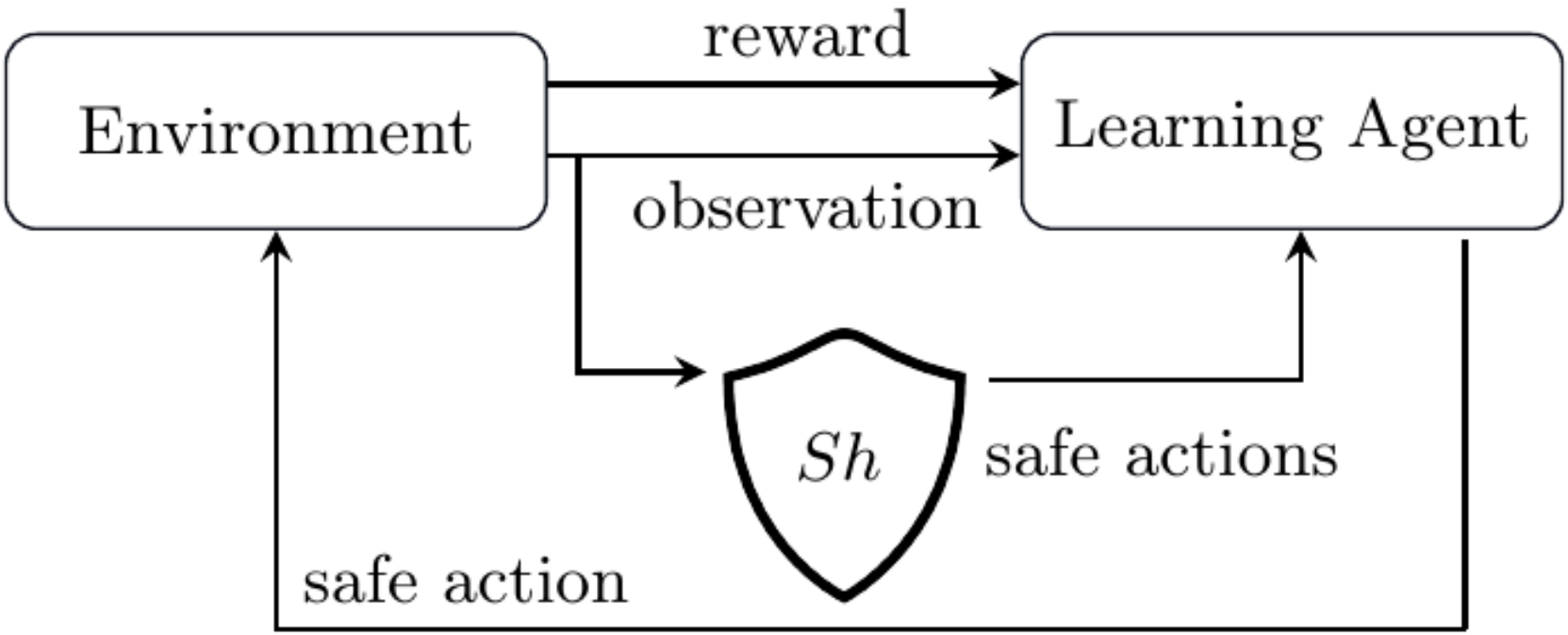
Post-shielded agent



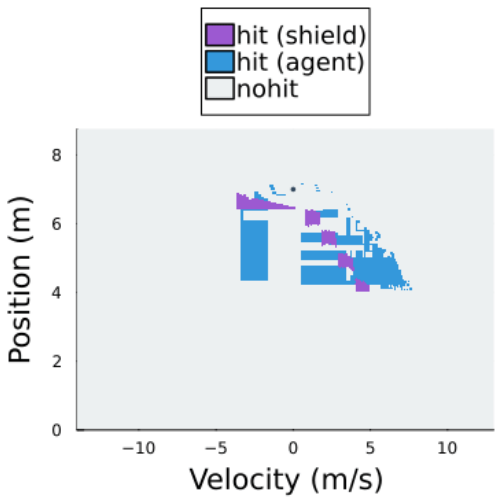


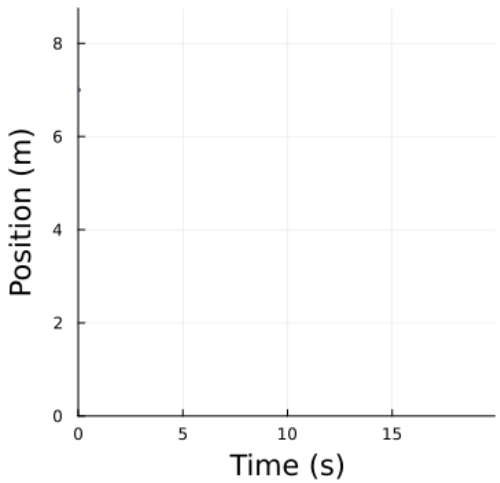


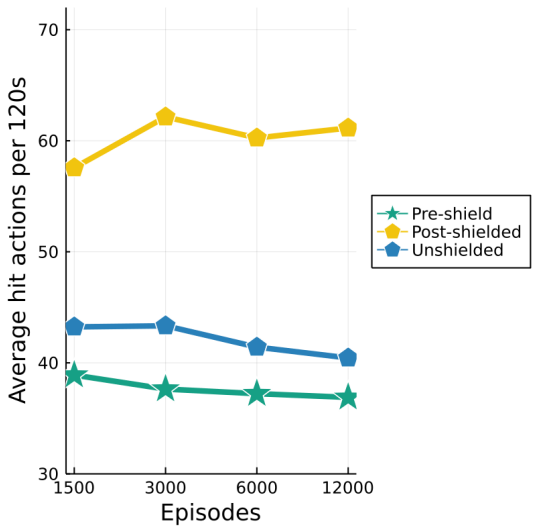
Pre-shielded agent



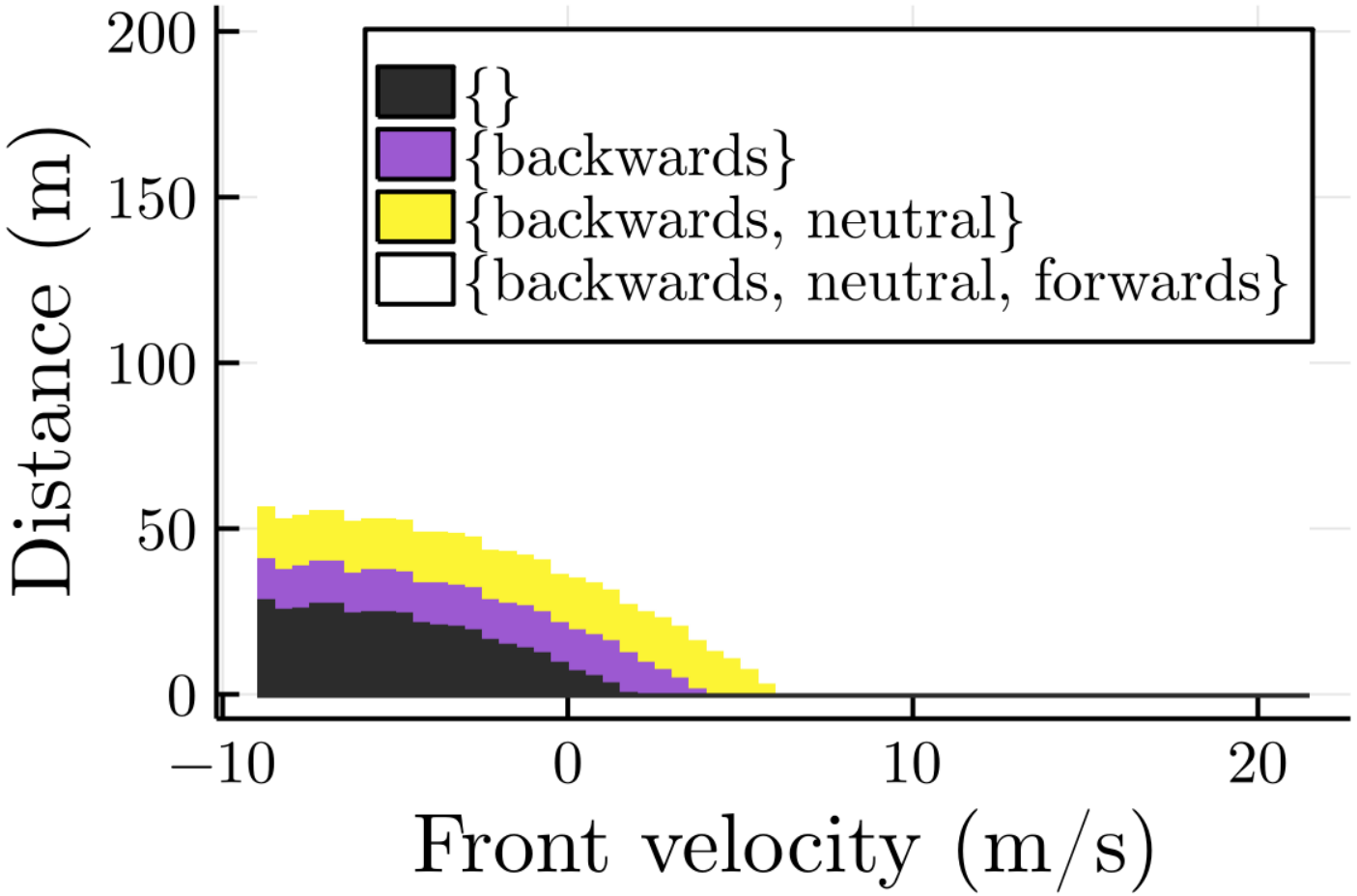
Pre-shielded agent





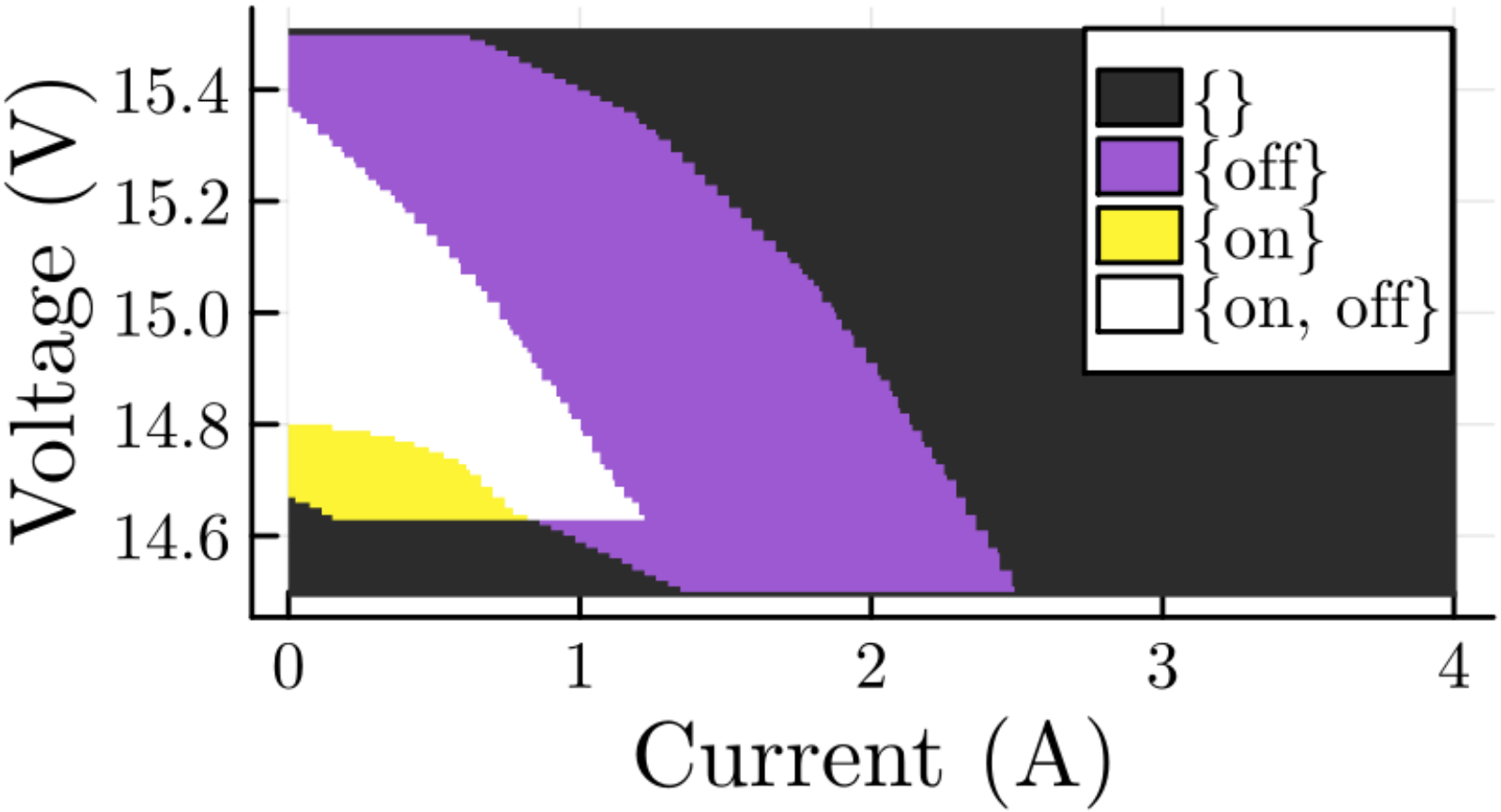


Cruise control (car behind another car)



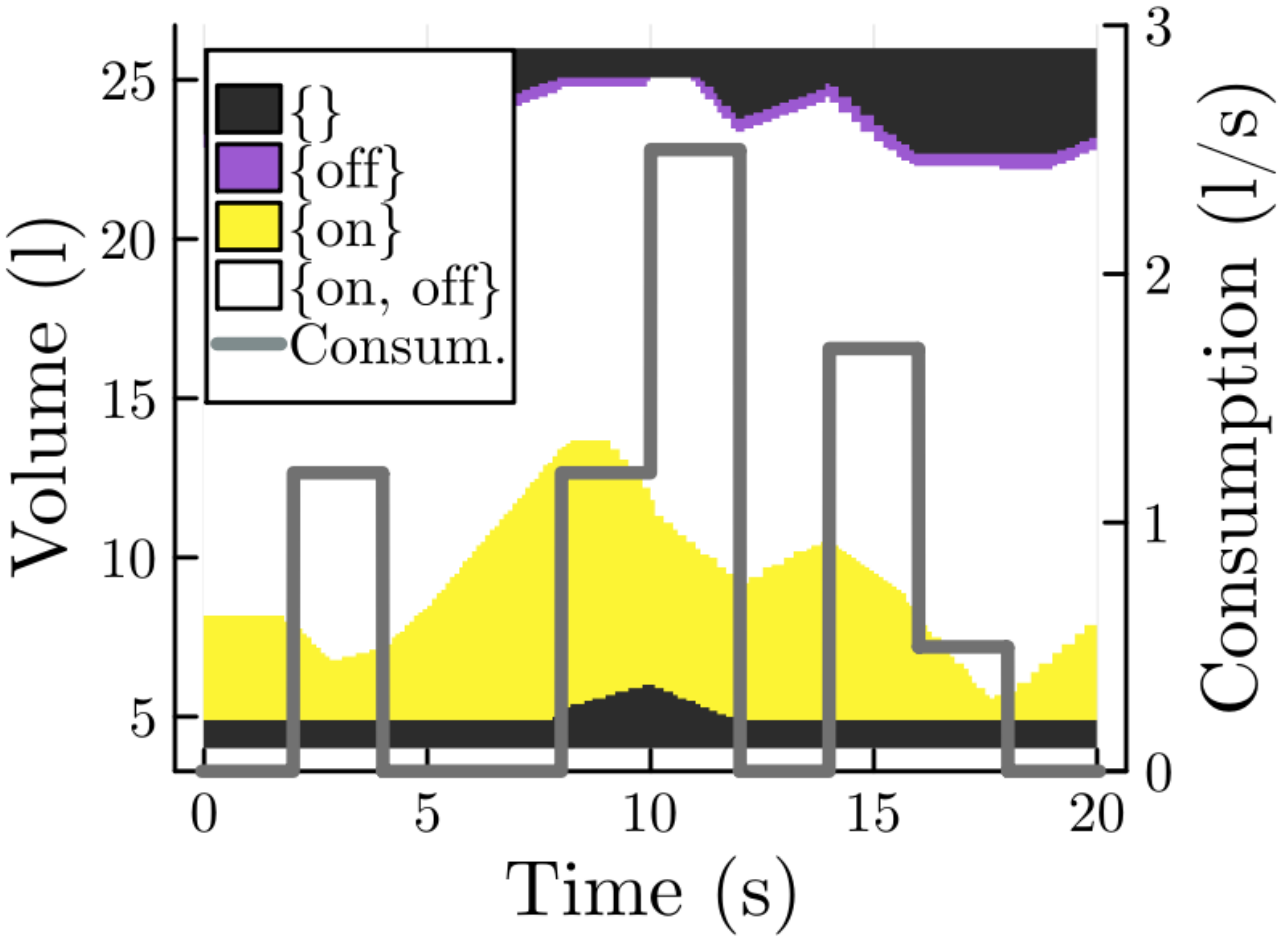
Synthesis time: 36 min

DC-to-DC boost converter



Synthesis time: 1 h 19 min

Oil pump (here: strategy when pump is *on*)



Synthesis time: 5 h 23 min

Overview

Conclusion

- Shield synthesis for complex (continuous + discrete + stochastic) systems
- Key idea: replace undecidable step (reachability) by simulation
- No soundness guarantee, but statistically safe

Future work

- Multiple steps for more permissive shield
- Dynamic partitioning
- Combine with symbolic approach