Shield Synthesis for AI

Bettina Könighofer
Roderick Bloem
TU Graz

Thomas Henzinger
Guy Avni
Krishnendu Chatterjee
IST AUSTRIA
Institute of Science and Technology

Ufuk Topcu
Scott Niekum
Mohammed Alshiek
Suda Bharadwaj
TEXAS
The University of Texas at Austin

Nils Jansen
IN·DEI·NOMINE·FELICITER
Radboud Universiteit Nijmegen

Rüdiger Ehlers
Universität Bremen

Sebastian Junges
RWTH AACHEN UNIVERSITY

Rayna Dimitrova
UNIVERSITY OF LEICESTER

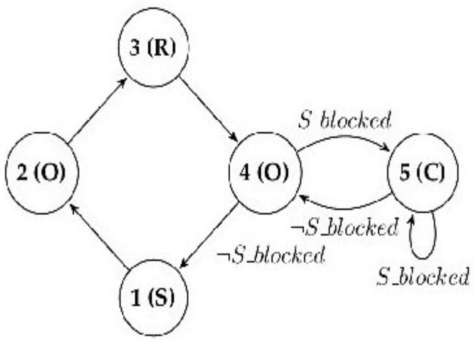
Reactive Synthesis | **Reinforcement Learning**

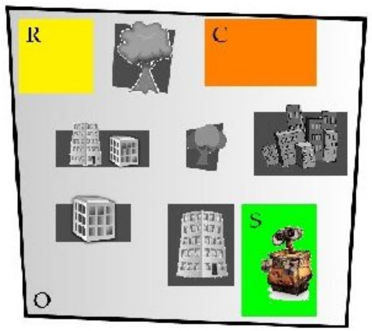
$C \quad \models \quad \phi$

Synthesis

via Game Solving

Your Controller



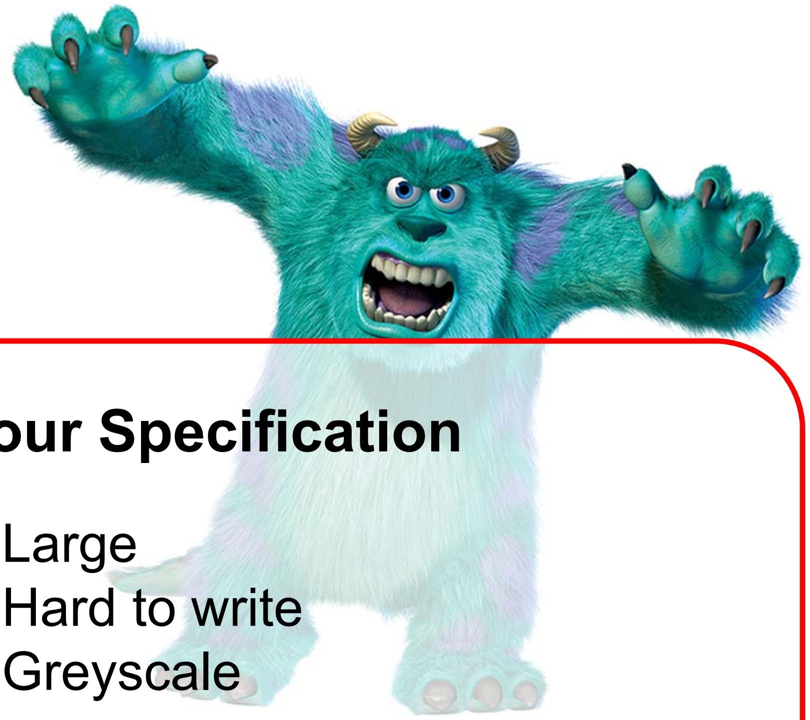


Your Specification

Infinitely often, visit R and S.
If S is blocked, go to C. Resume visiting R and S once S is unblocked.

$$\mathbf{G}(\neg blocked \rightarrow \mathbf{F} R) \wedge \mathbf{G}(\neg blocked \rightarrow \mathbf{F} S) \wedge \mathbf{G}(blocked \rightarrow \mathbf{X}(C\ \mathbf{U}\ \neg blocked))$$

Reactive Synthesis **Reinforcement Learning**

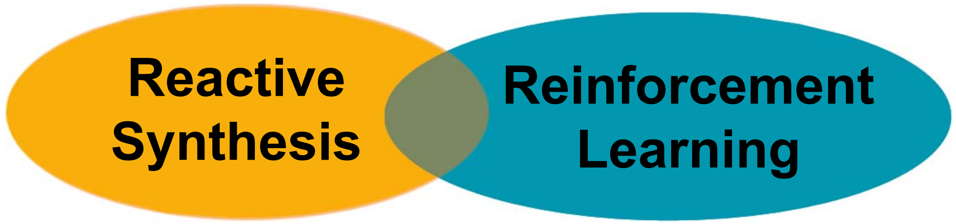


Your Controller

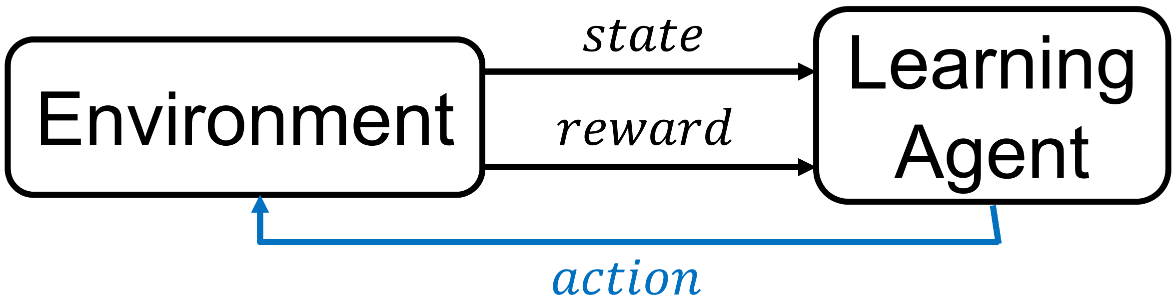
- Large
- Complicated
- Highly optimized
- Many sensors
- …

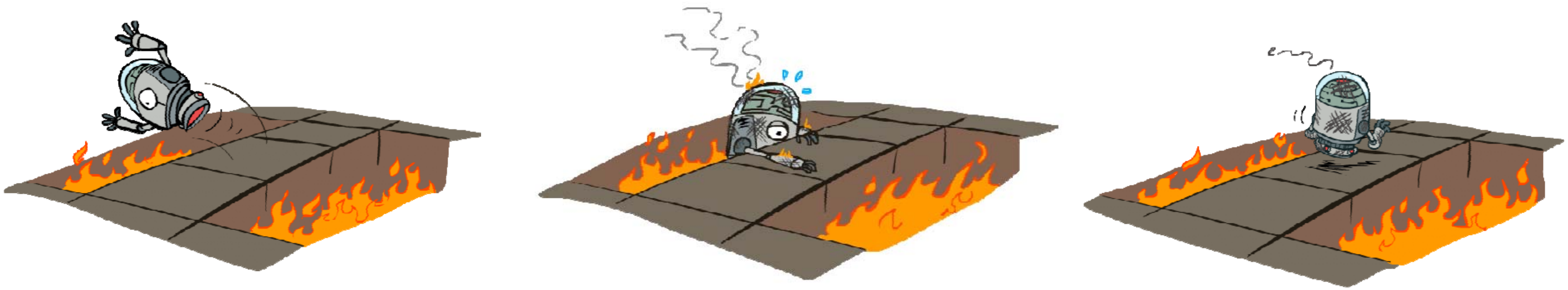
Your Specification

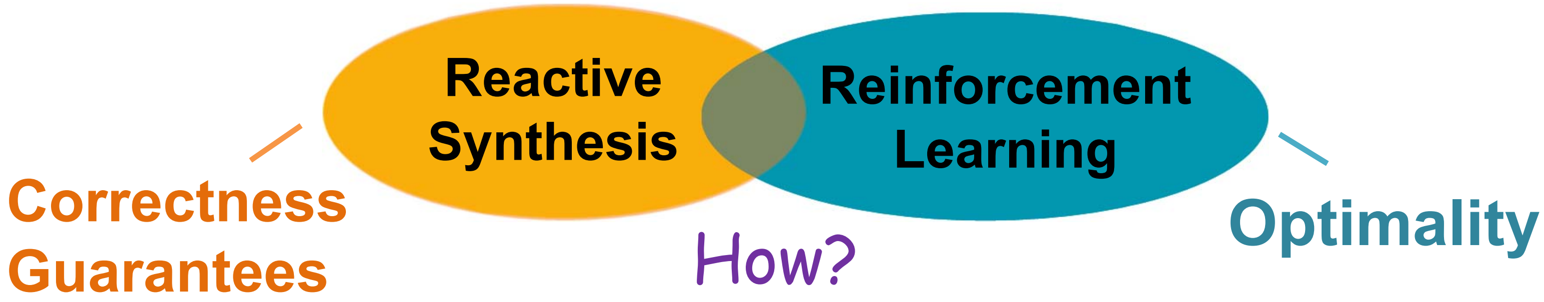
- Large
- Hard to write
- Greyscale



Reinforcement Learning






Reactive Synthesis
Reinforcement Learning
Correctness Guarantees
How?
Optimality

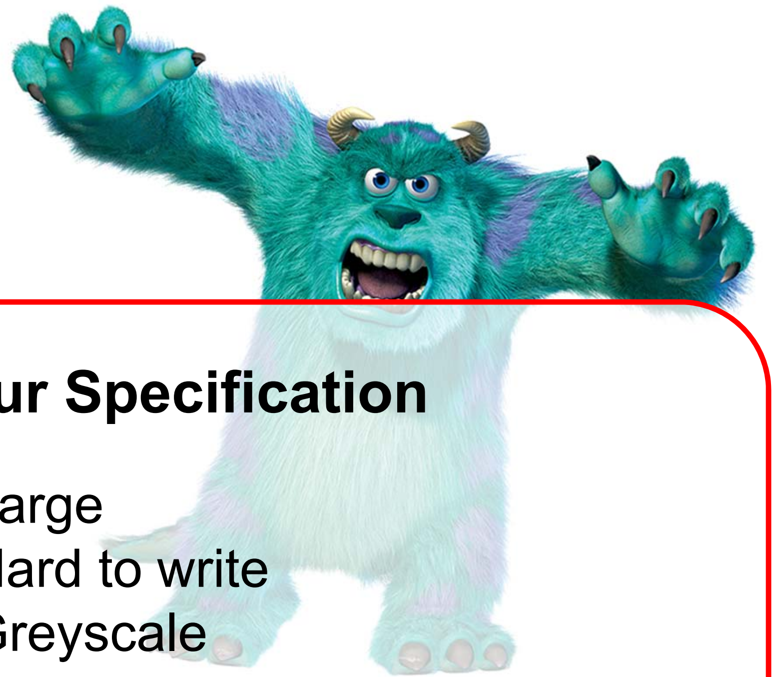
Reactive Synthesis

Reinforcement Learning

Correctness Guarantees

Optimality

Shielding



Your Controller
- Large
- Complicated
- Highly optimized
- Many sensors
- …

Your Specification
- Large
- Hard to write
- Greyscale

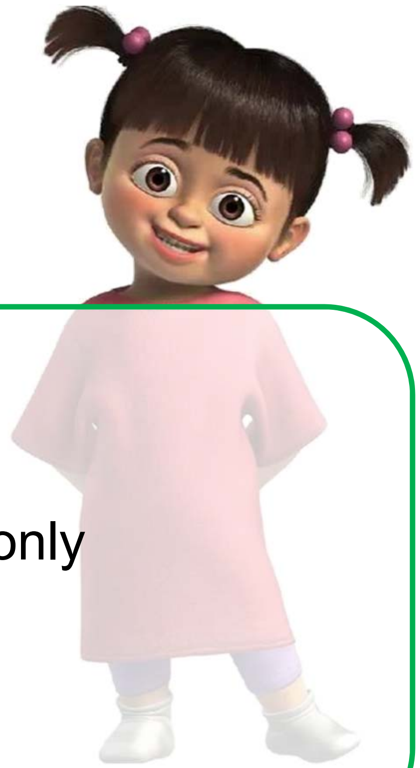
Reactive Synthesis

Reinforcement Learning

Correctness Guarantees

Optimality

Shielding

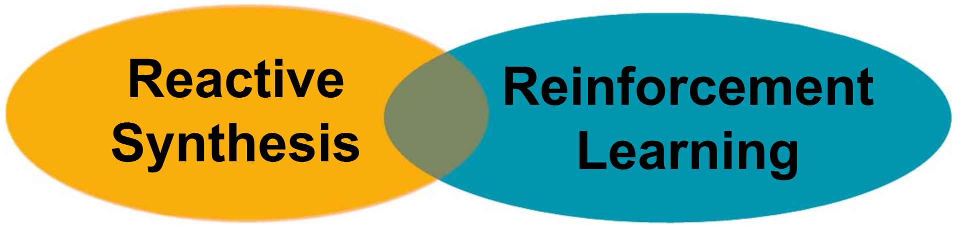


Your Controller

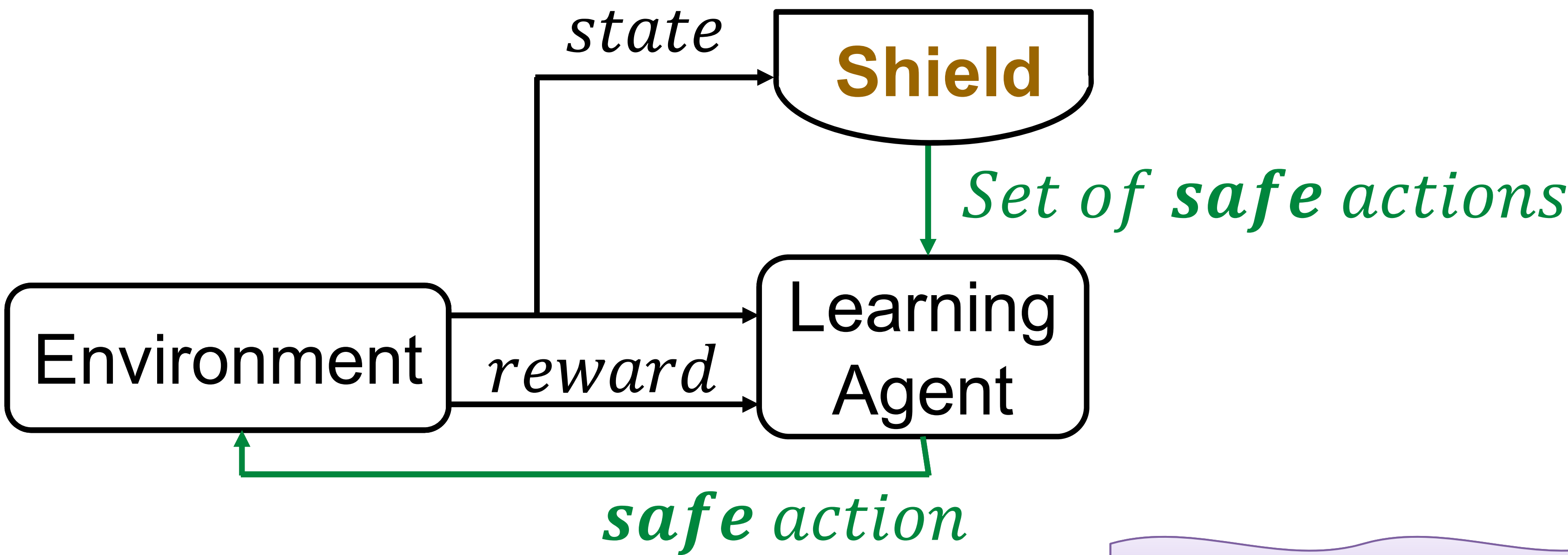
- Large
- Complicated
- Highly optimized
- Many sensors
- …

Critical Spec

- Critical aspects only
- Small & sweet

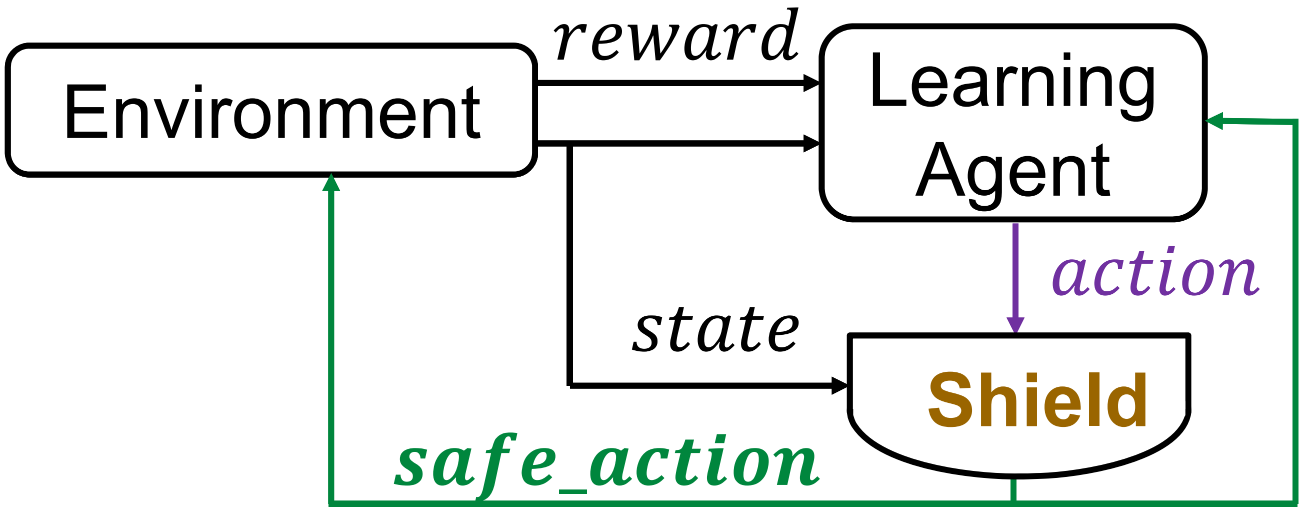


Preemptive Shielding



Minimal Interference

Post-Posed Shielding



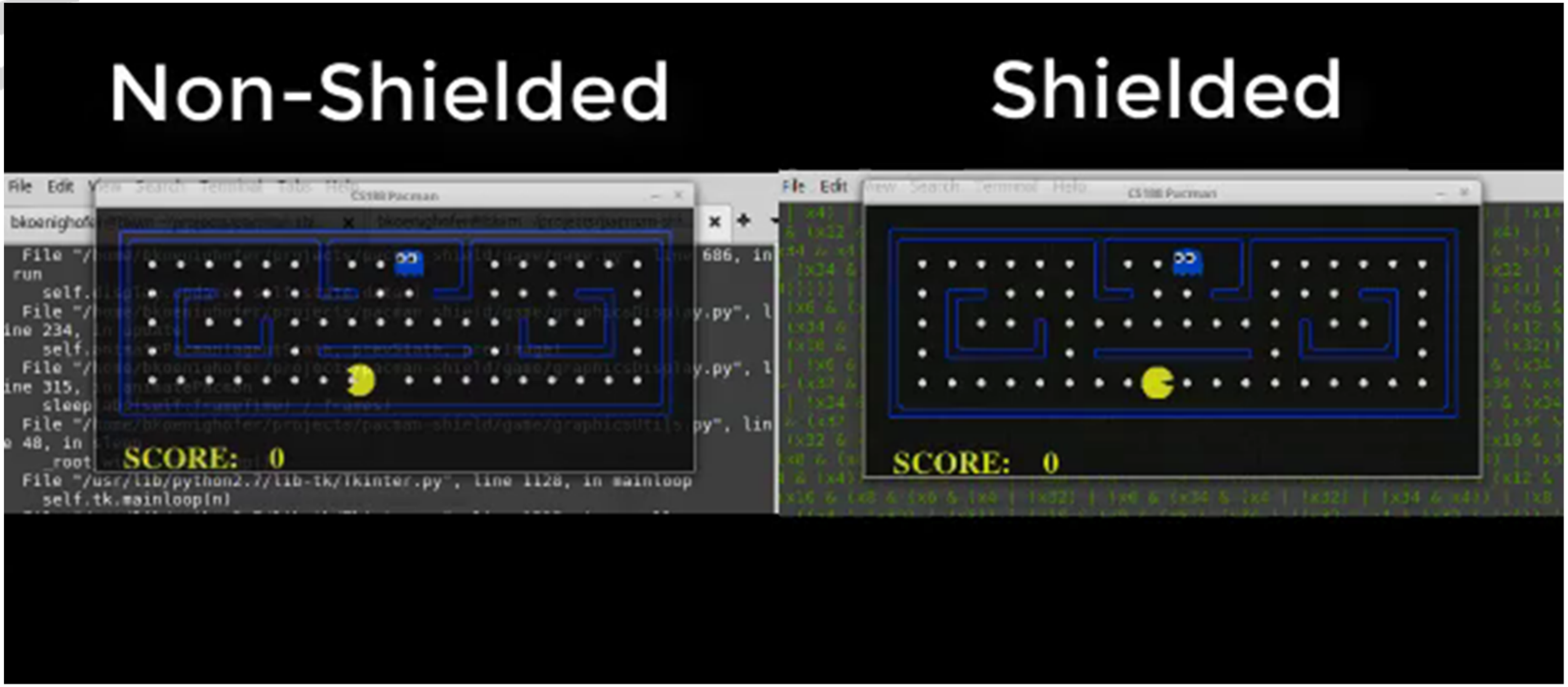
Policy Update:

- for $safe\_action$ using $reward$
- for $action$ if $action \neq safe\_action$:
  1. Assign a punishment to $action$
  2. Assign $reward$ to $action$

Shield can be added in execution phase

A Shield for PAC-MAN

M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, U. Topcu:
Safe Reinforcement Learning via Shielding. AAAI 2018



Outline

- Safety Shields AAAI-18 ✓
- Optimal Shields submission
- Safety Shields for Multi-Agent Systems ACC-19
- Probabilistic Safety Shields arXiv

Optimal Shields

- Problems of learned controllers
  - (Safety problems)

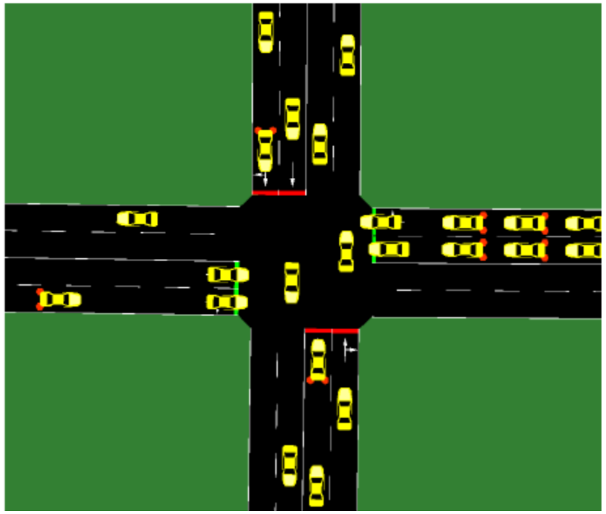
1. Difficult to add **new features**
2. **Poor performance** on **un-trained** behavior
3. No local fairness

Solution:
Optimal Shield

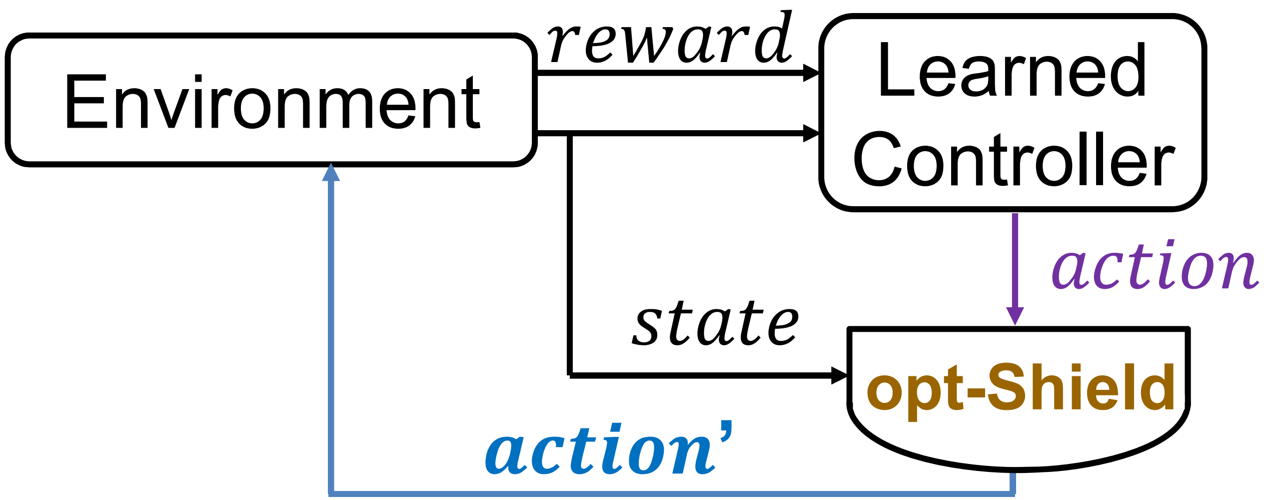
Shields for Traffic Light Controllers

- Learned Controller: “minimize total waiting time”

1. Difficult to add **new features**
   - priority to public transport, changes due to an accident
2. **Poor performance** on **un-trained** behavior
   - Uniform traffic congestion meets rush-hour traffic
3. No local fairness
   - Farm road never gets green



Optimal Shields Synthesis



- Lightweight shields → Two cost functions
  - c_{BEH}: Cost for behavior
  - c_{INT}: Cost for interference

$$\lambda \cdot c_{BEH} + (1 - \lambda) \cdot c_{INT}$$

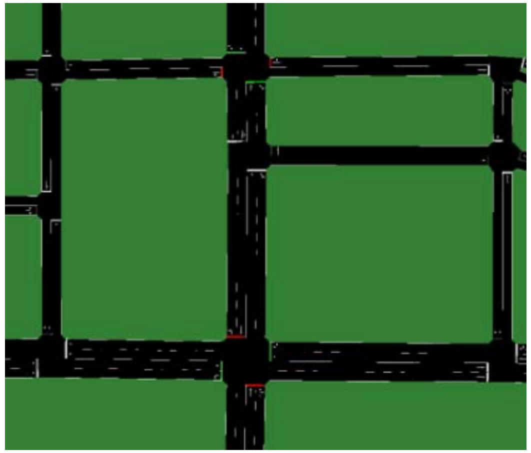
Mean-Payoff Game with 2 Objectives

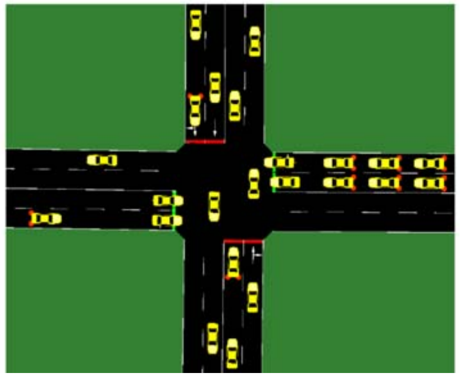
Mean-Payoff Game

λ: tradeoff between objective of controller vs shield

Dealing with rush-hour traffic

- Controller
  - Deep Convolutional Q-Network
    - 16 dim input vector
    - num approaching cars, waiting time
    - 4 layers (16, 604, 604, 4 nodes),
    - Q-learning: $\alpha = 0.001, \gamma = 0.95$
  - “Minimize waiting time of two junctions”
- Shield
  - c_{BEH}: size of maximal queue
  - c_{INT}: 1 for interference, 0 otherwise

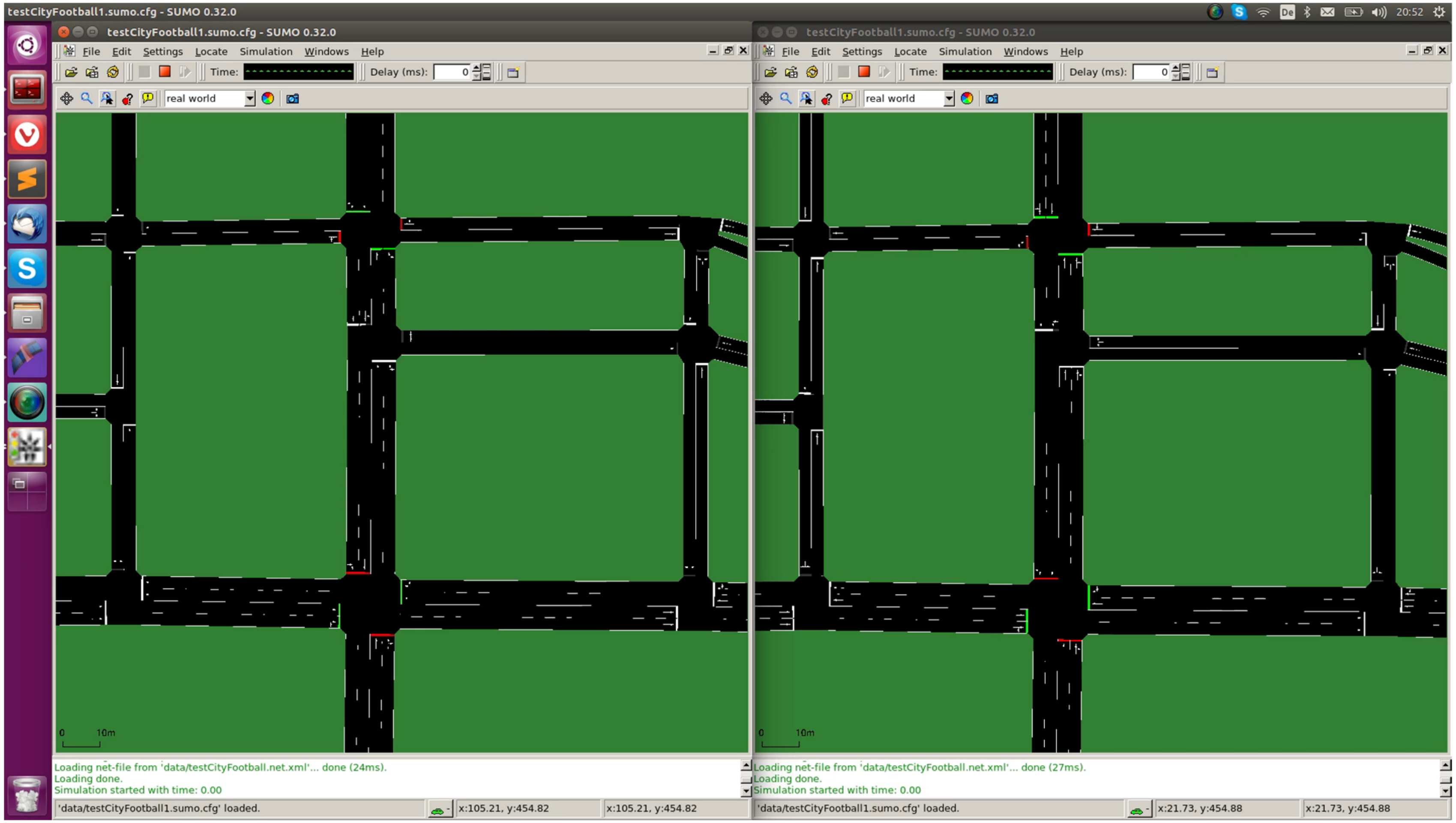




abstract state
(1,8,1,2)

Dealing with rush-hour traffic

G. Avni, B. Könighofer, T. Henzinger, K. Chatterjee, R. Bloem:
Run-Time Optimization for Learned Controllers through Quantitative Games. Under submission.



Outline

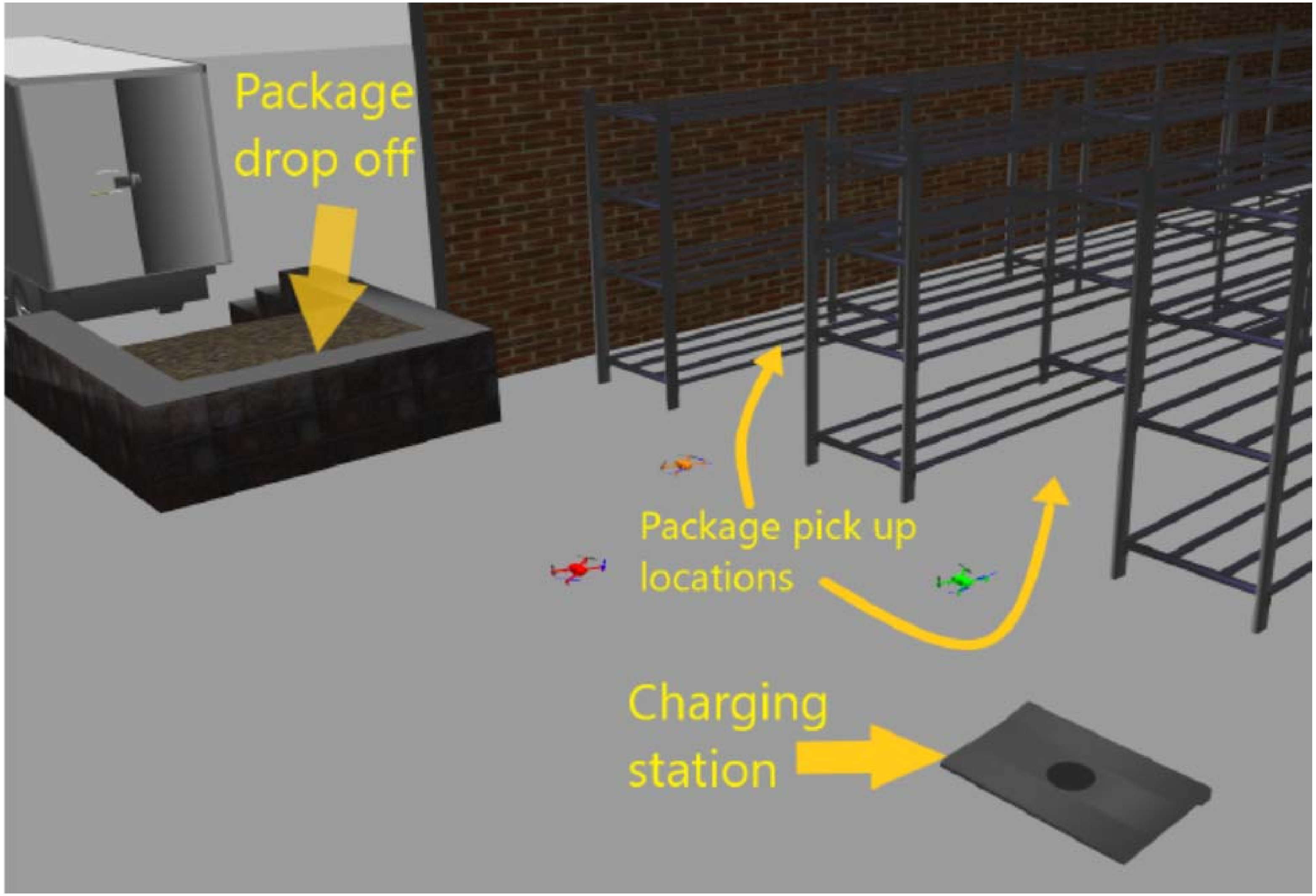
- Safety Shields AAAI-18 ✓
- Optimal Shields submission ✓
- Safety Shields for Multi-Agent Systems ACC-19
- Probabilistic Safety Shields arXiv

Safety Shields for Multi-Agent Systems

- Task: Enforce global safety property

1. Quantitative interference costs c_{INT}:
   - Counting cost function
   - Different costs for interferences with different agents
2. Fair Shielding
   - Do not always interfere with the same agent repeatedly

Case Study: Warehouse



Case Study: Warehouse

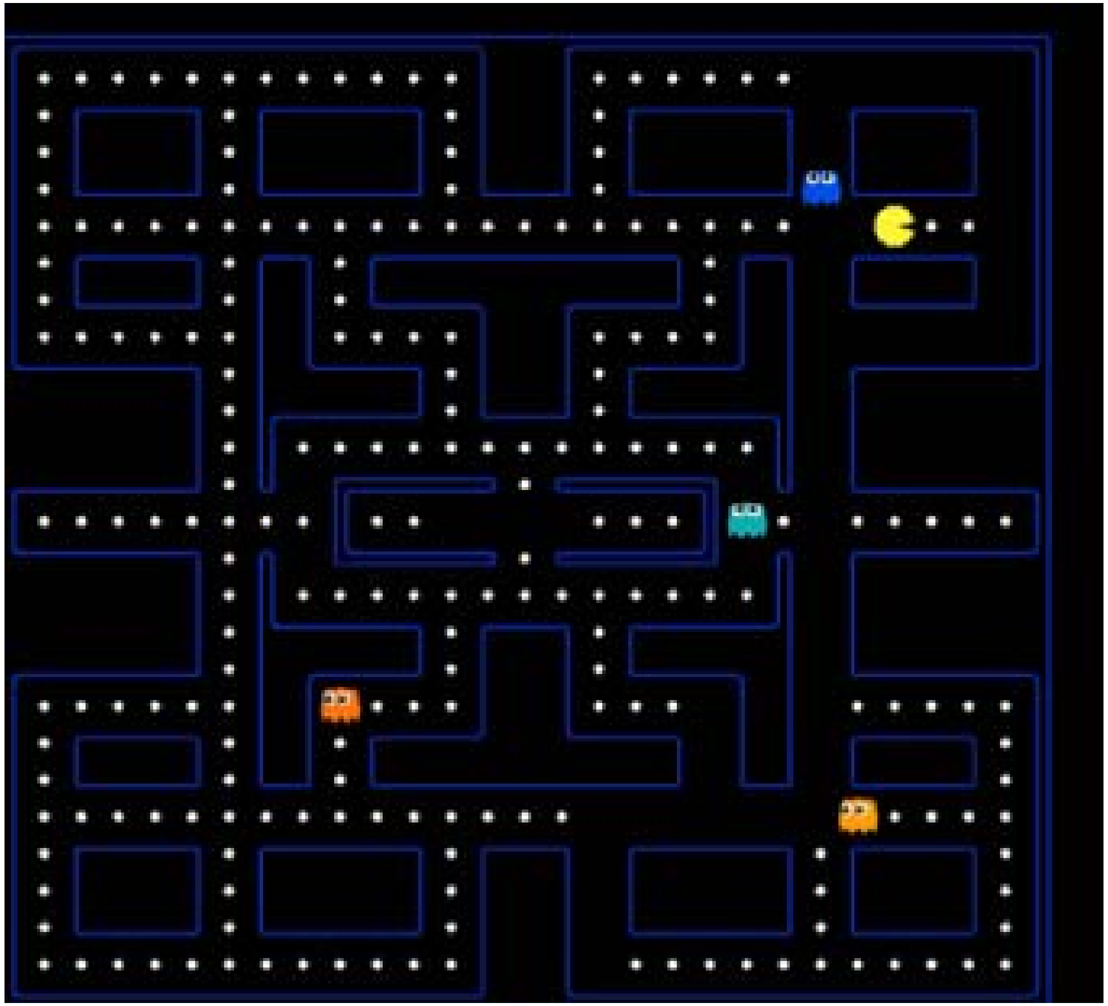
S. Bharadwaj, R. Bloem, R. Dimitrova, B. Könighofer, and U. Topcu: Synthesis of Minimum-Cost Shields for Multi-agent Systems. ACC-19

Outline

- Safety Shields AAAI-18 ✓
- Optimal Shields submission ✓
- Safety Shields for Multi-Agent Systems ACC-19 ✓
- Probabilistic Safety Shields arXiv

Shielding original Pacman?

- State space is huge!
- Not realizable!



Learning the Adversary Model

- Each ghost has it‘s individual behaviour
  - Observe it, model the behaviour
  - Data augmentation techniques
    - Is PAC-MAN north, south, east, or west?

- Results in MDP of environment
- Guaranteed safety w.r.t. **probabilistic** temporal logic spec

MDP is huge! Scalability

- Finite Horizon
  - safety for finite number of steps
  - infinite horizon may cause large errors anyways

- Piecewise Construction
  - compute shield for each state independently

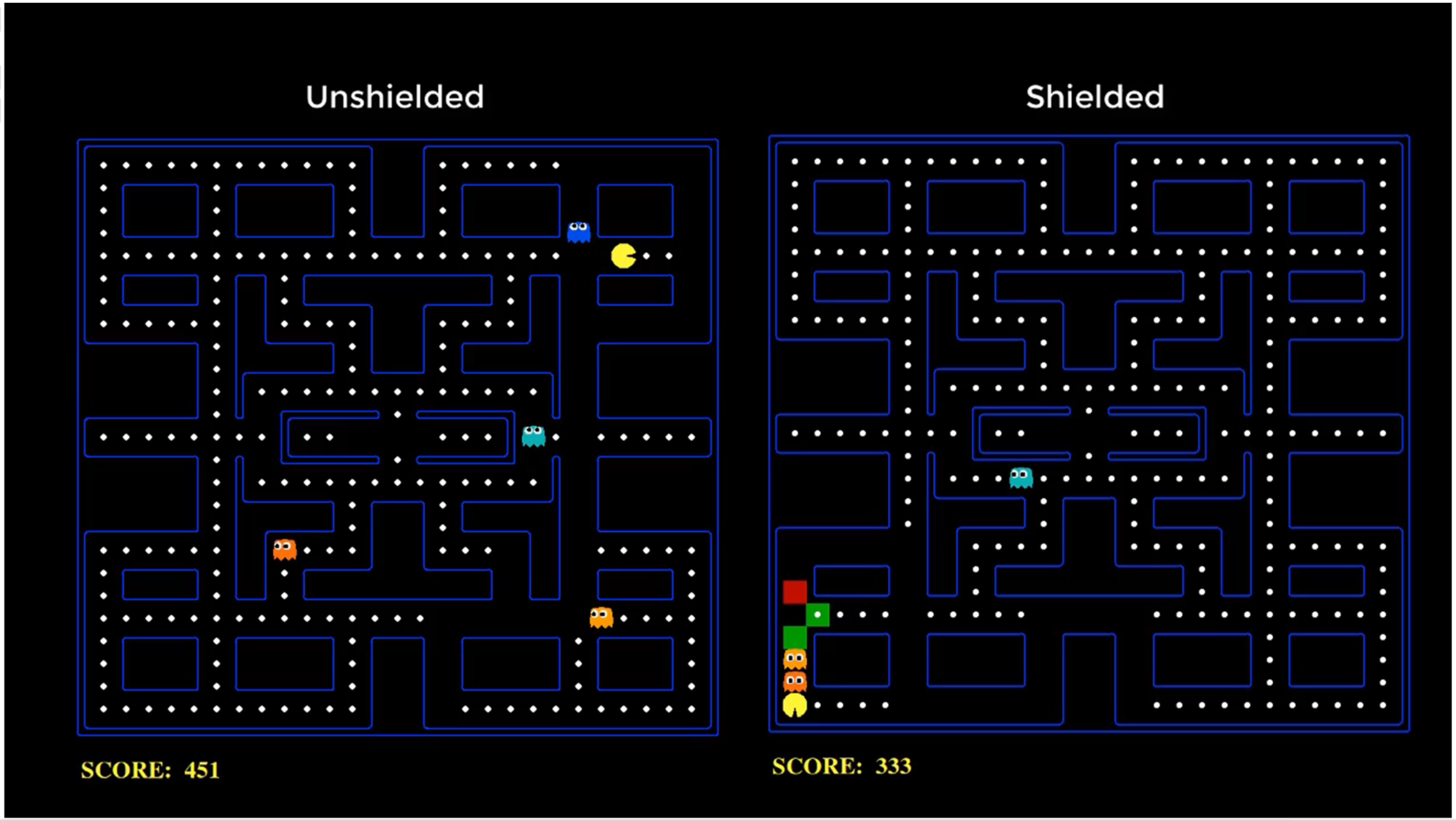
MDP is huge! Scalability

- Independent Agents
  - crashing probabilities for different agents are stochastically independent
  - compute individually, compose shields

- Abstractions
  - adversaries may be far away
  - neglect adversary positions that are not relevant

Probabilistic Safety Shield for Pacman

N. Jansen, B. Könighofer2, S. Junges, and R. Bloem:
Shielded Decision-Making in MDPs, arXiv



Future Work

- Safety Shields
  - Shields for CPS, Deal with wrong models
- Optimal Shields
  - Performance in autonomous systems
- Safety Shields for Multi-Agent Systems
  - Distributed Shield Synthesis
- Probabilistic Safety Shields
  - Partially observable MDPs

THANK YOU!