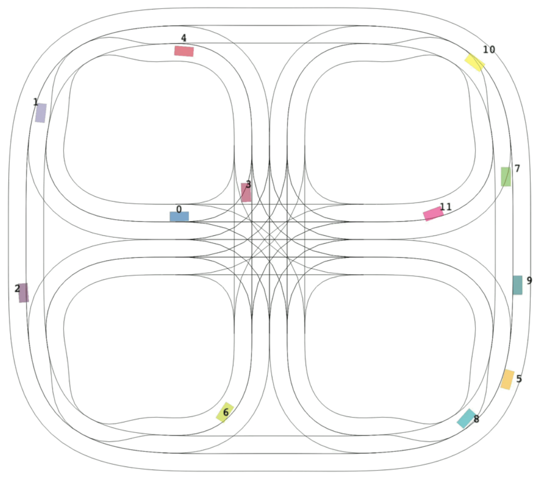
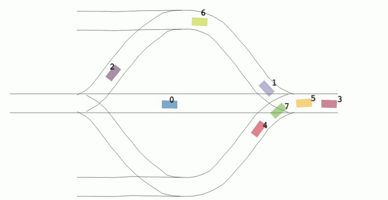
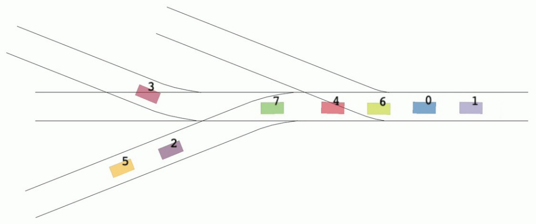
ITSC 2024 | Edmonton, Canada

Presenter: Simon Schäfer

September 23, 2024







SigmaRL: A Sample-Efficient and Generalizable Multi-Agent Reinforcement Learning Framework for Motion Planning

Jianye Xu, Pan Hu, Bassam Alrifaee

Informatik 11 Embedded Software | RWTH AACHEN UNIVERSITY

Universität der Bundeswehr München

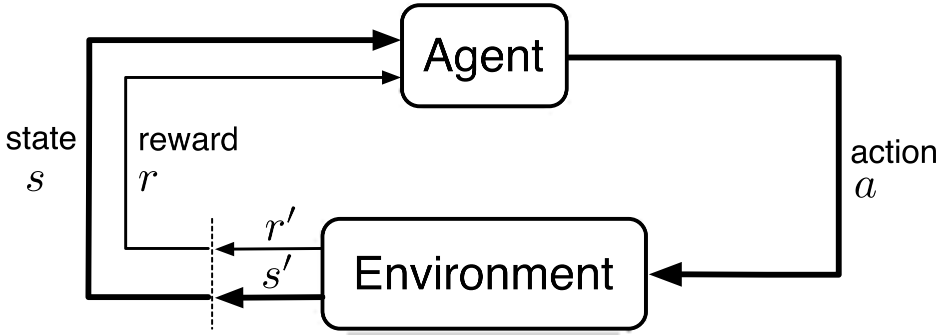
Preprint | Code | Video

Introduction

► Multi-Agent Reinforcement Learning (MARL) for motion planning of Connected and Automated Vehicles (CAVs)

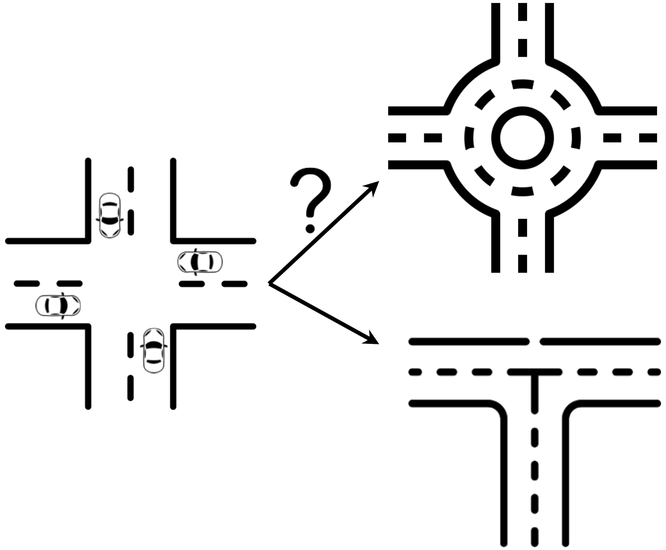
1. Sample efficiency

A *sample* $:= (s, a, r', s')$



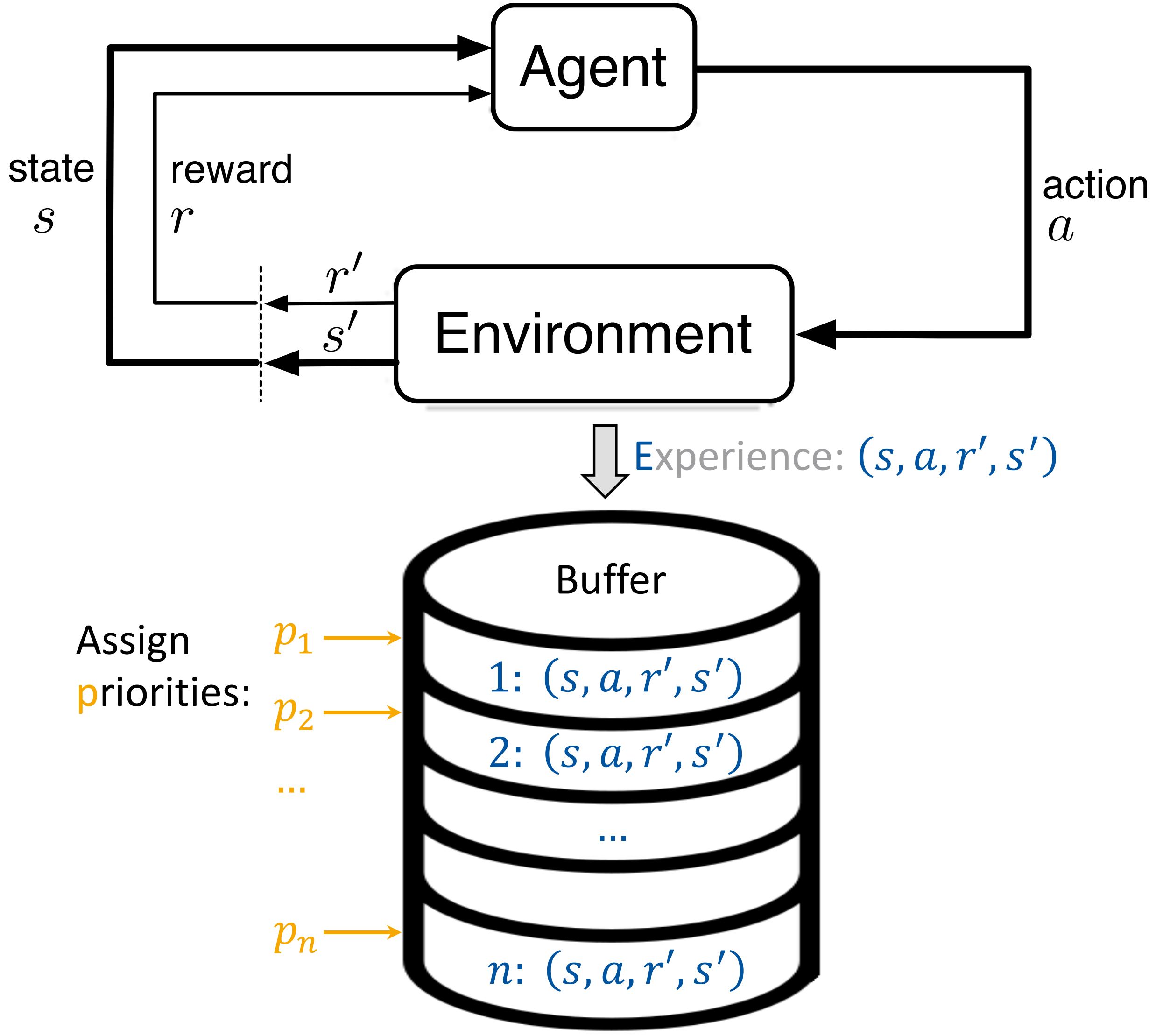
R. S. Sutton et al., Fig. 3.1 modified

2. Generalization



Related Work | Enhance Sample Efficiency

- Experience replay buffer [1]
  - Independent and identically distributed (i.i.d.) assumption
- Prioritized experience replay buffer [2]



[1] L.-J. Lin, "Self-Improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching," *Mach Learn*, 1992.
[2] T. Schaul et al., "Prioritized Experience Replay," *in International Conference on Learning Representations (ICLR)*, 2016.

Related Work | Enhance Generalization

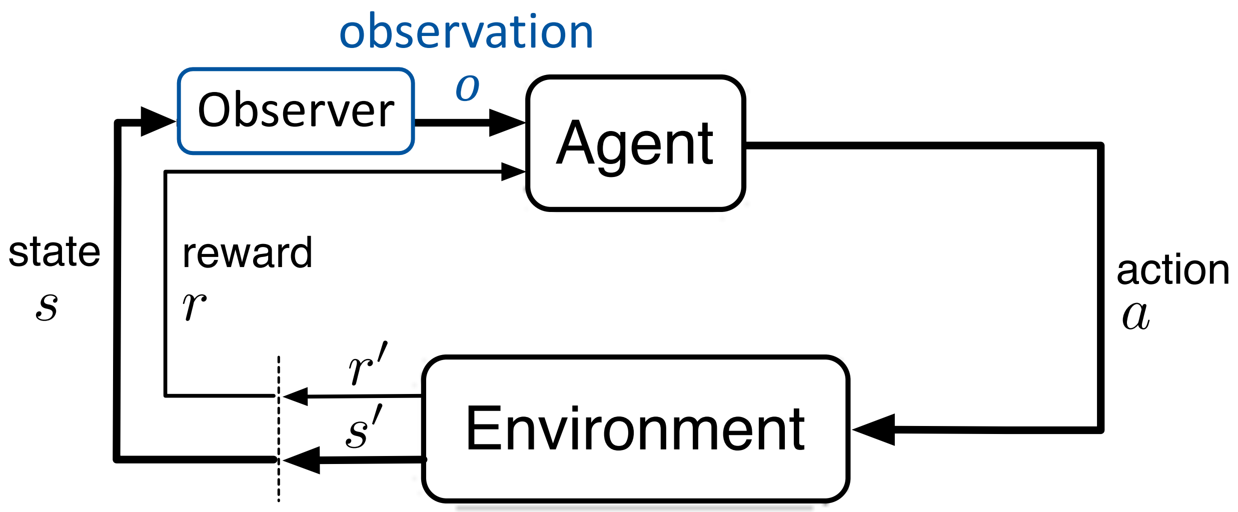
- Regularization techniques such as dropout and early stopping [1]
- Domain randomization [2]
- Data augmentation [3]
- Better optimization without overfitting [4]

[1] J. Farebrother et al., “Generalization and Regularization in DQN,” *arXiv*, 2020.
[2] Tobin, Josh, et al., “Domain randomization for transferring deep neural networks from simulation to the real world,” *IEEE/RSJ IROS*, 2017.
[3] Connor Shorten, and Taghi M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, 2019.
[4] Cobbe, Karl W., et al. "Phasic policy gradient." *International Conference on Machine Learning*, 2021.

Related Work | Gap

- Observation design: under-explored

- Observer (also called observation function)

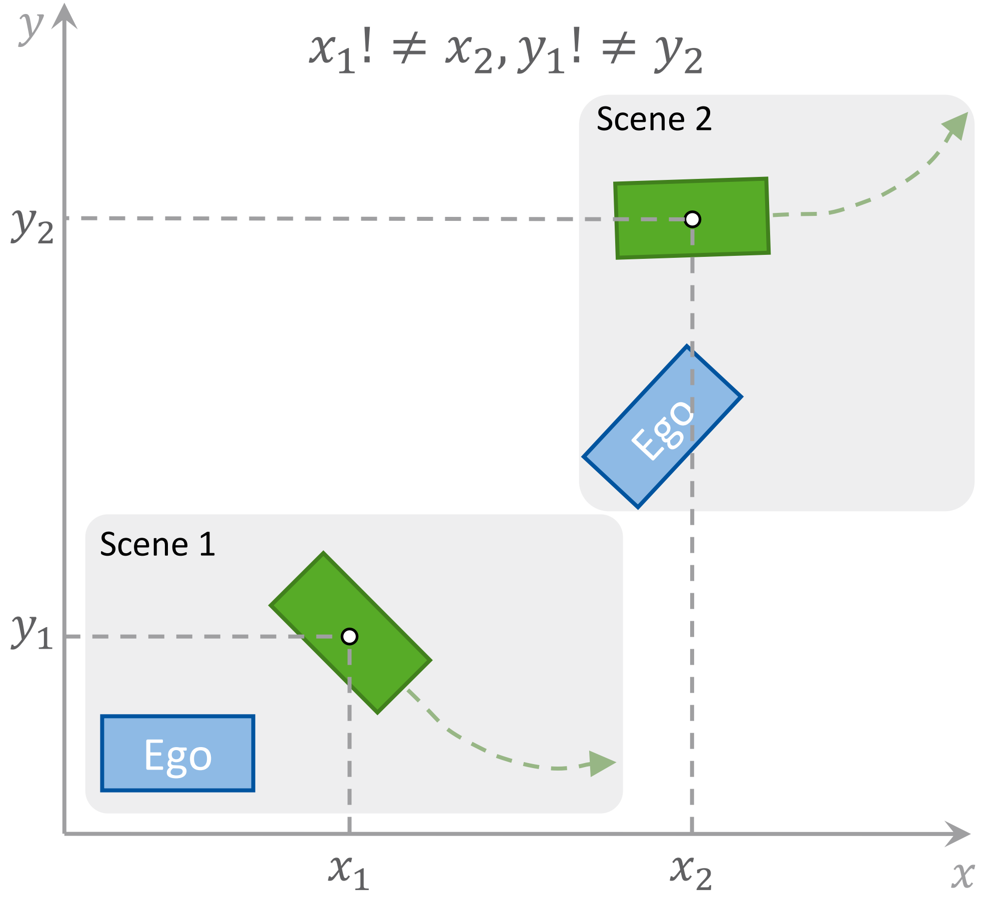


- Five observation-design strategies

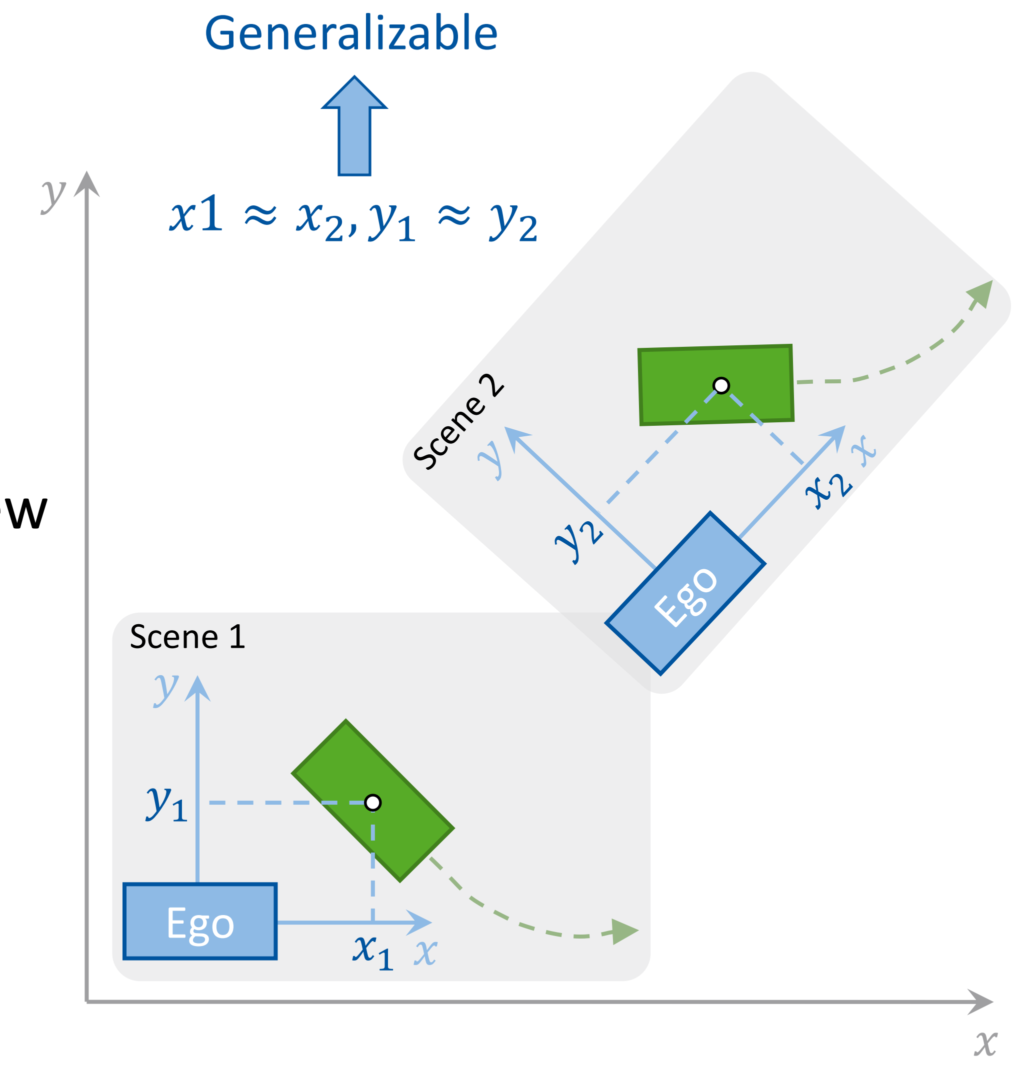
Methodology | Observation-Design Strategy 1/5

- Use **ego view** instead of bird-eye view

Bird-eye view:



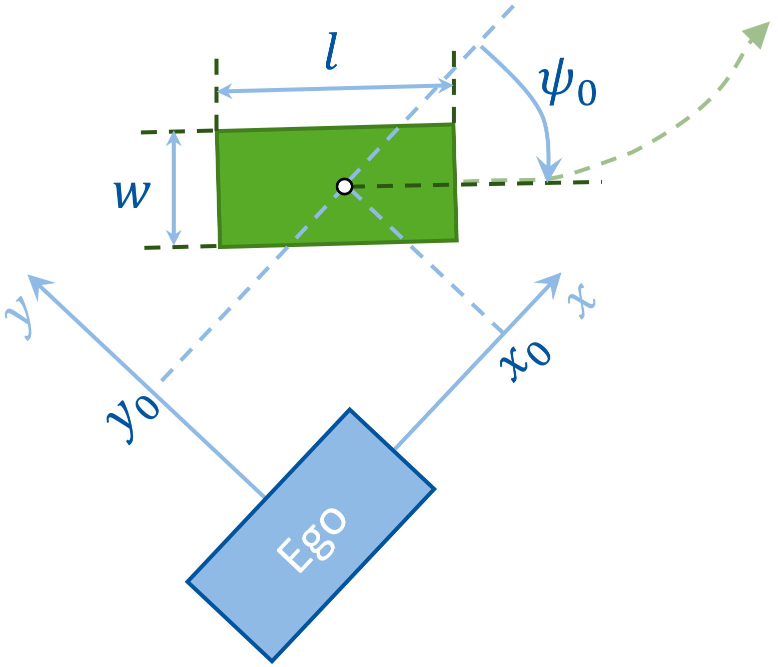
Ego View (**our**):



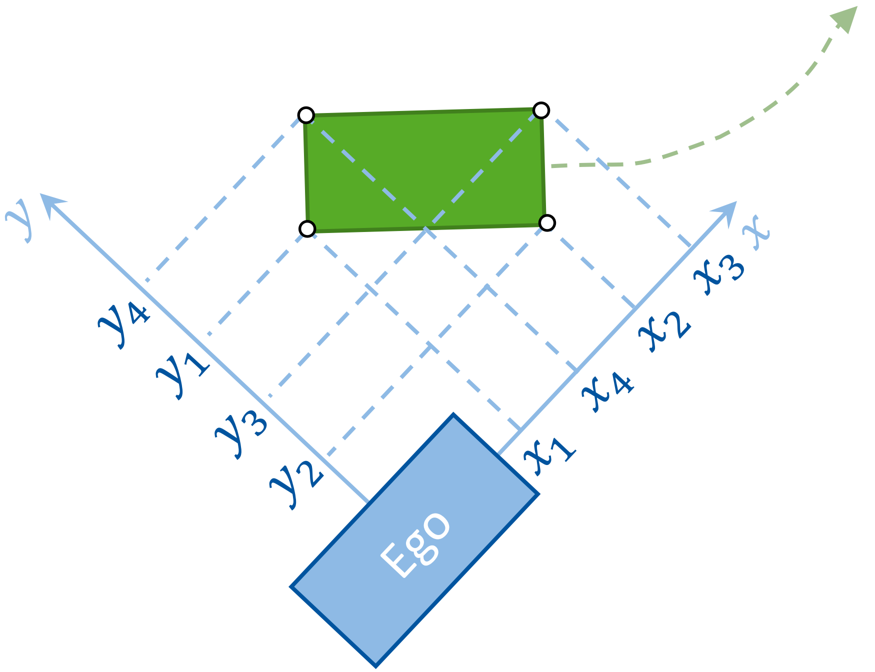
Methodology | Observation-Design Strategy 2/5

- Observe **vertices** of surrounding agents instead of poses and dimensions

Poses and dimensions:

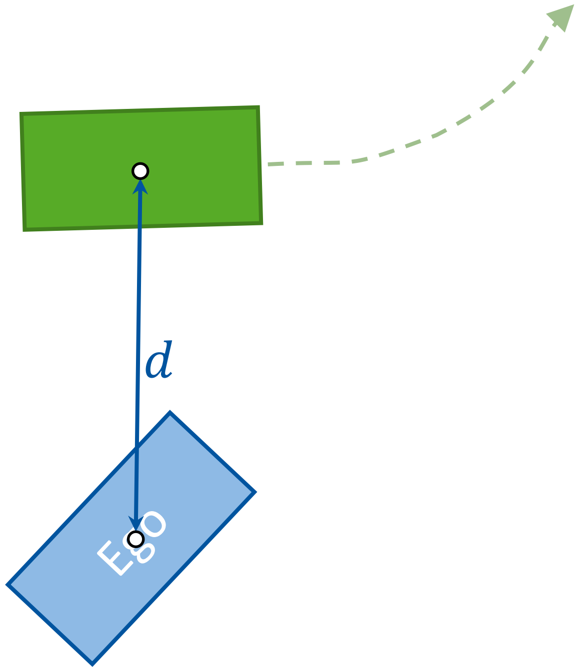


Vertices (**our**):



Methodology | Observation-Design Strategy 3/5

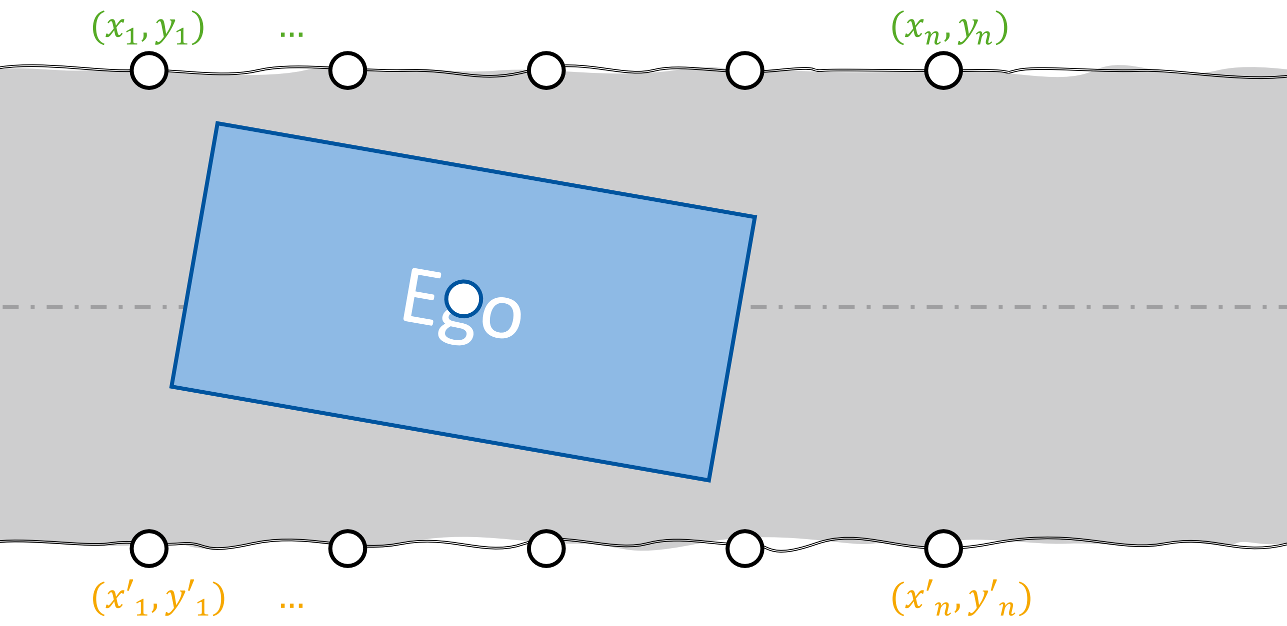
- Observe **distances** to surrounding agents



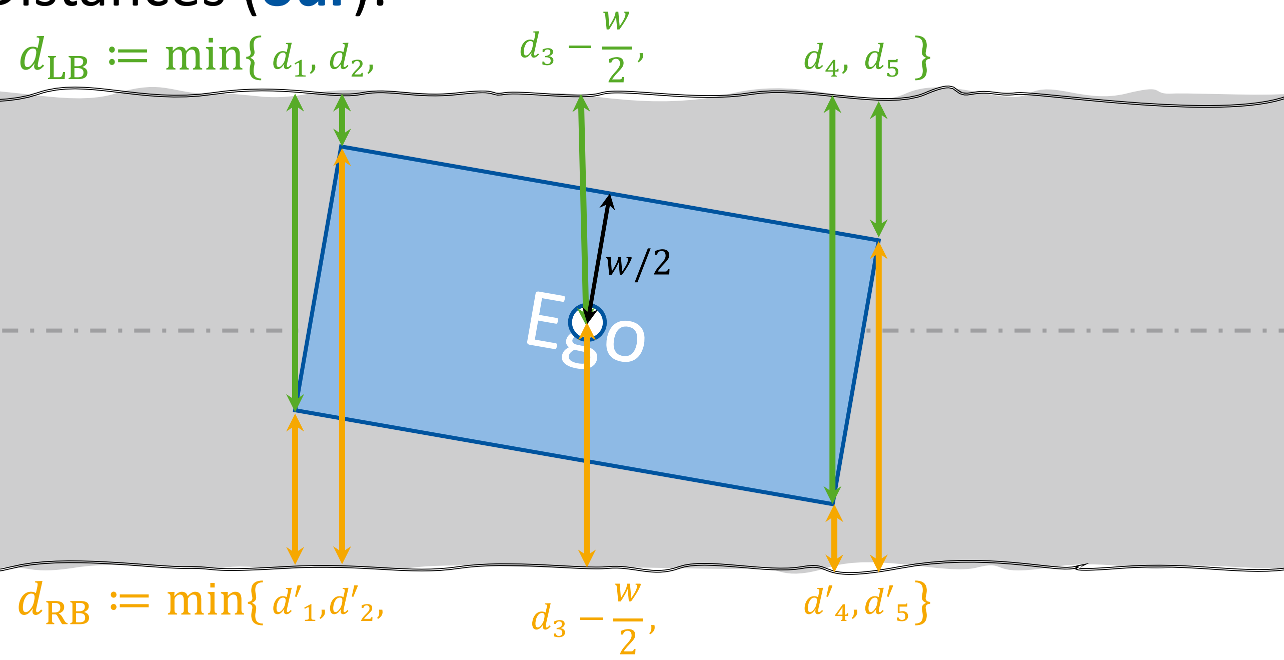
Methodology | Observation-Design Strategy 4/5

- Observe **distances** to lane boundaries instead of sampled points

Sample points:

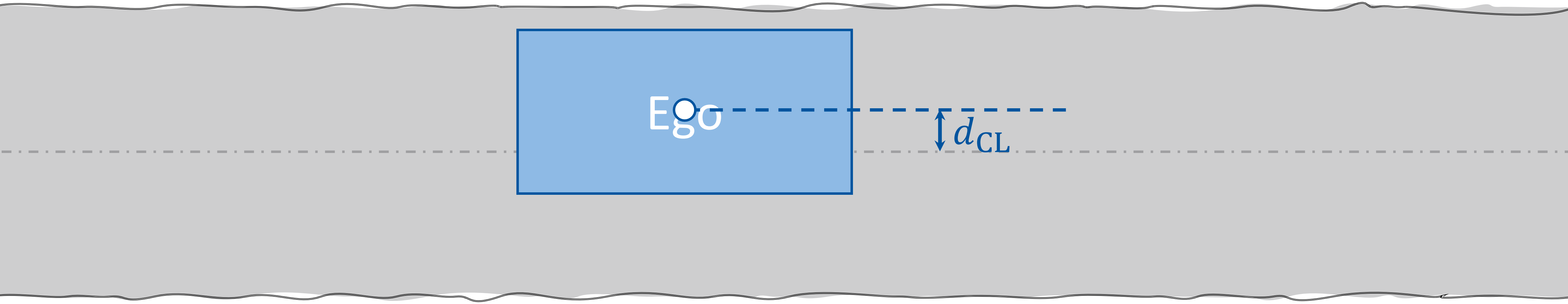


Distances (**our**):



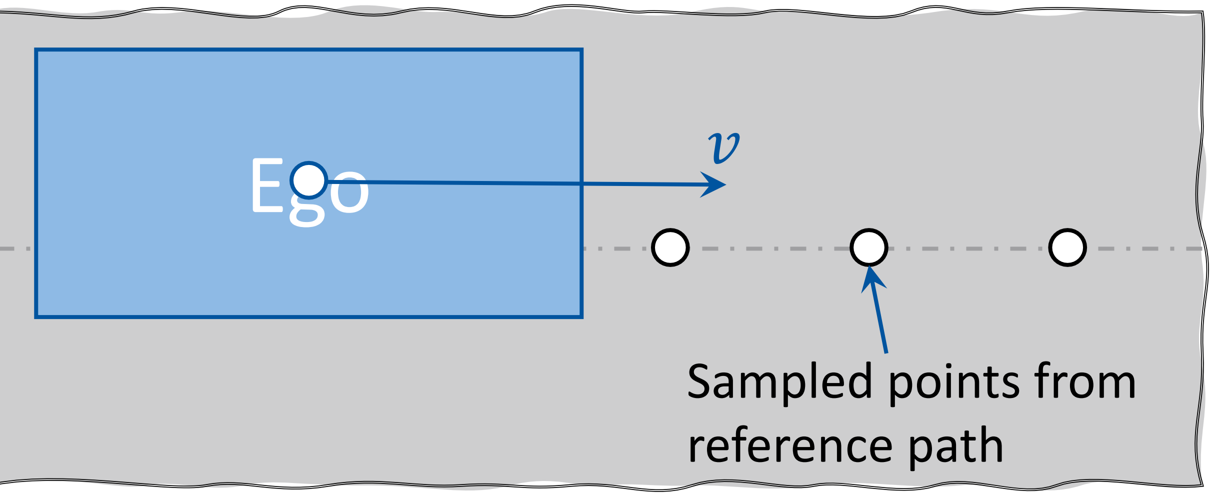
Methodology | Observation-Design Strategy 5/5

- Observe **distances** to lane center lines

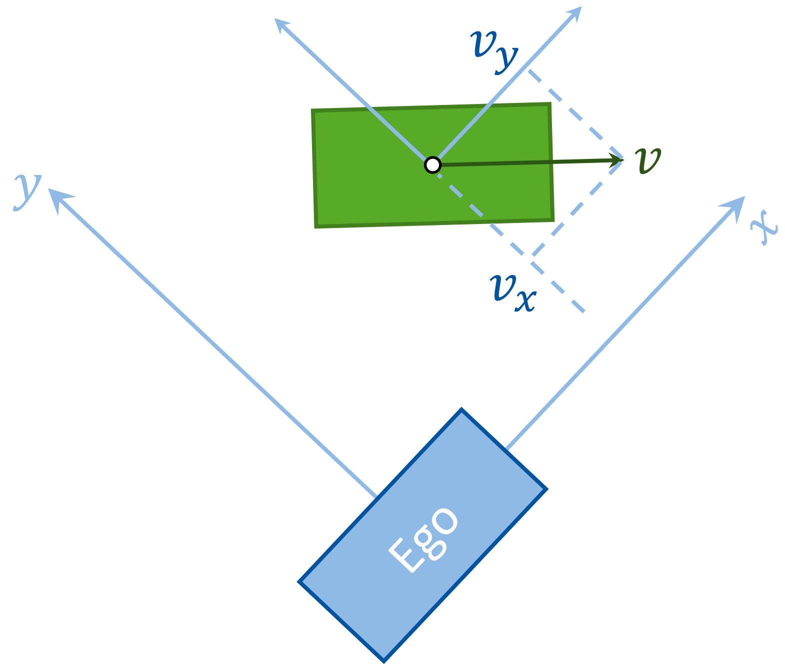


Methodology | Other Observations

- Own speed v
- Reference path

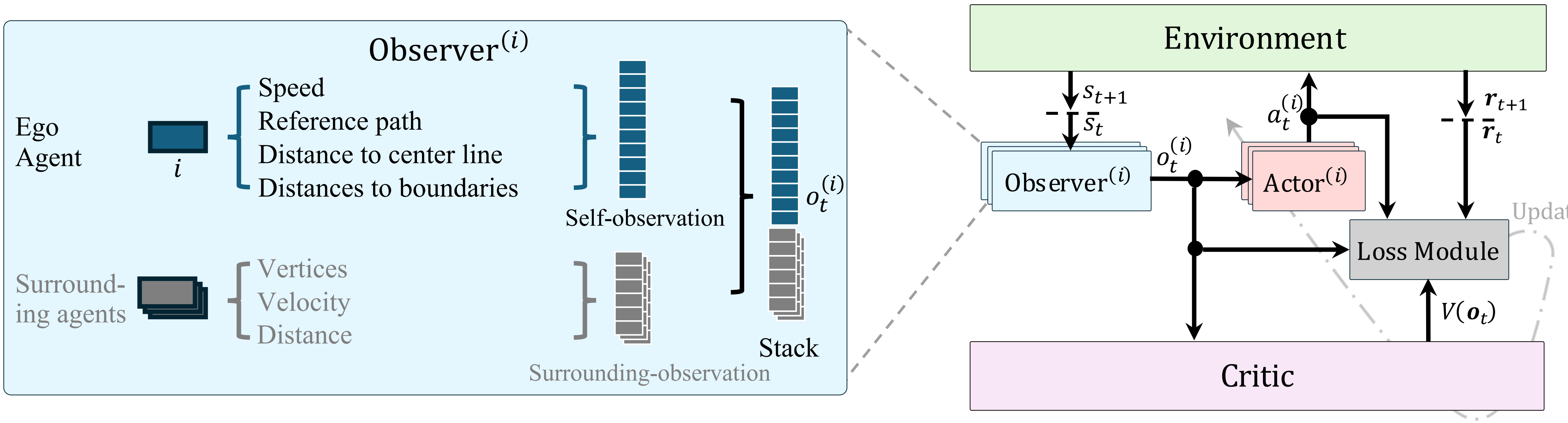


- Velocity of surrounding agents



Methodology | Framework Overview

- Integrated our observation-design strategies into multi-agent PPO [1-2]

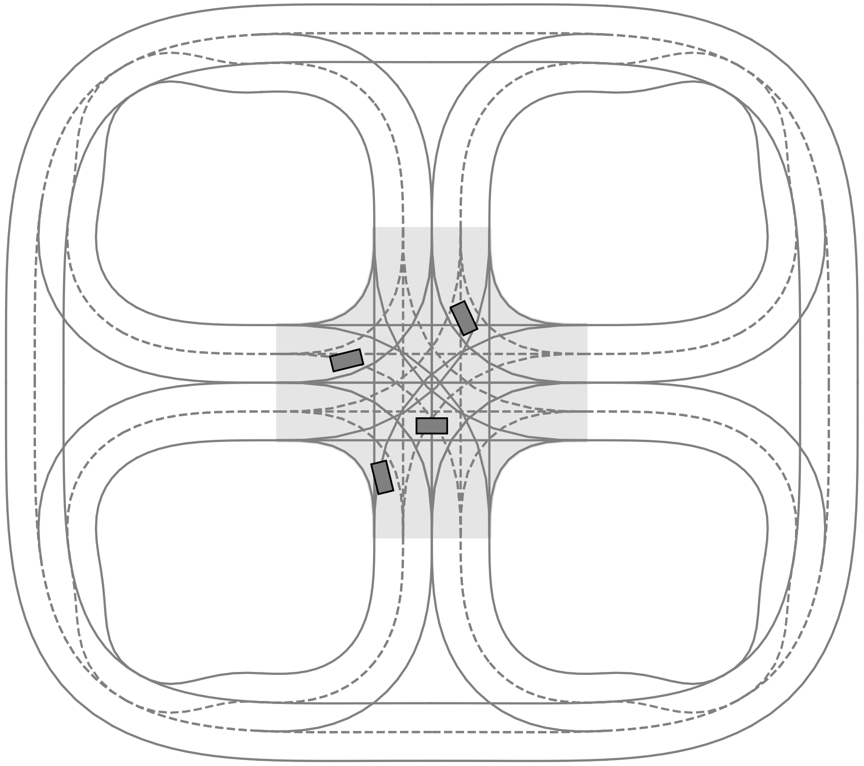


[1] J. Schulman *et al.*, "Proximal Policy Optimization Algorithms," *arXiv*, 2017.
[2] R. Lowe *et al.*, "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017

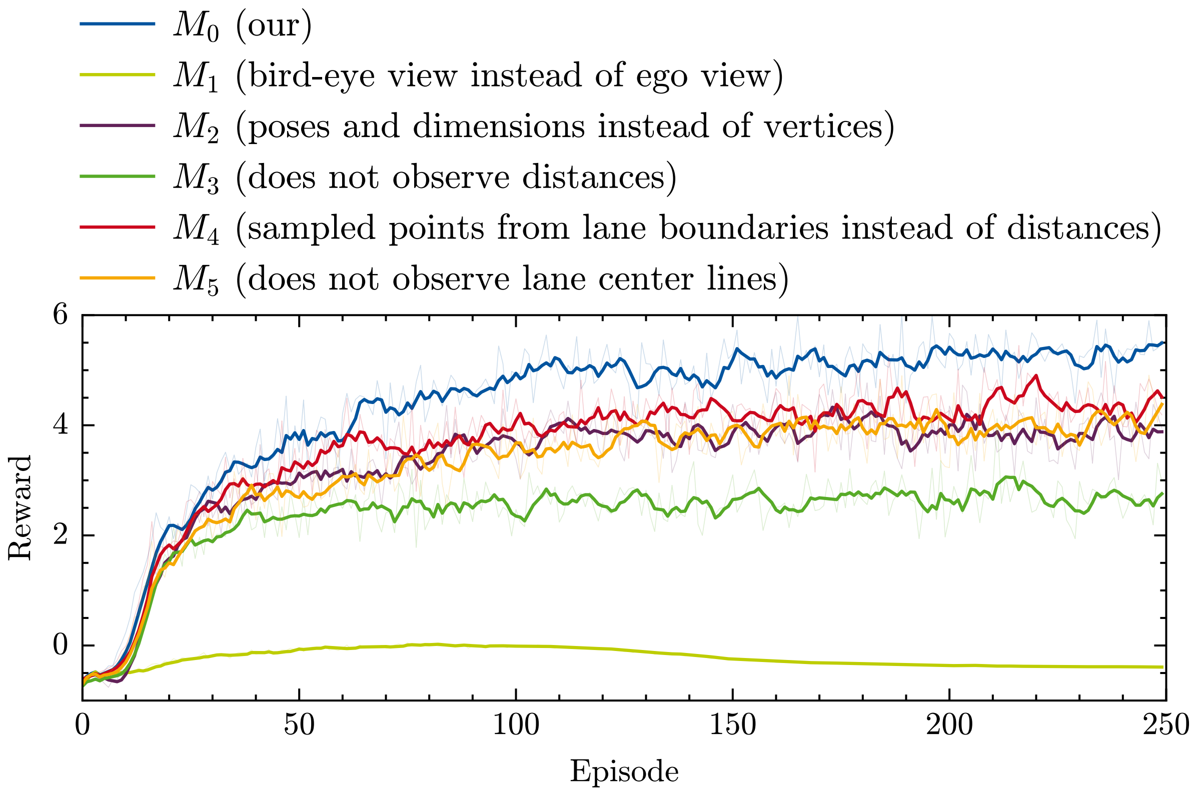
https://github.com/cas-lab-munich/sigmarl

Evaluations | Training

- An intersection with four agents
- ~1 million samples
- $<$ 1 h of training time with a single CPU

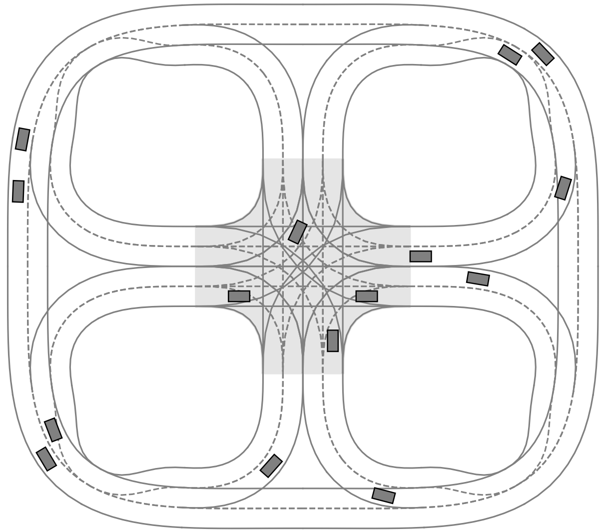


- Ablation studies with six models
  - M_0: All five observation-design strategies
  - $M_{i\in\{1,\ldots,5\}}$: Omits the i_{th} strategy

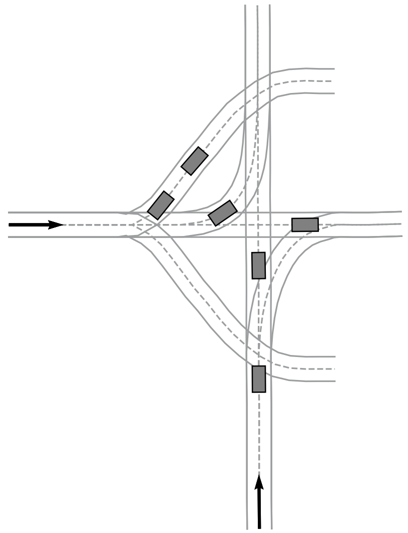


Evaluations | Testing Scenarios

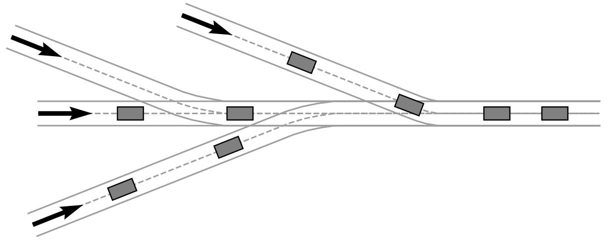
- Four unseen scenarios
- 32 one-minute simulations per scenario per model
- Collision rate of each simulation: proportion of time steps in which collisions occur



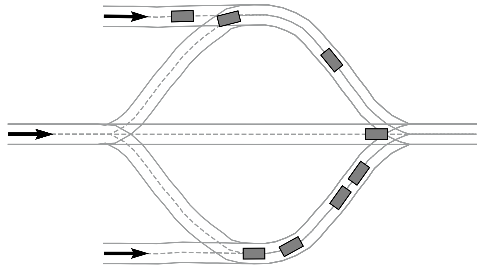
(a) With 15 agents



(b) With 6 agents

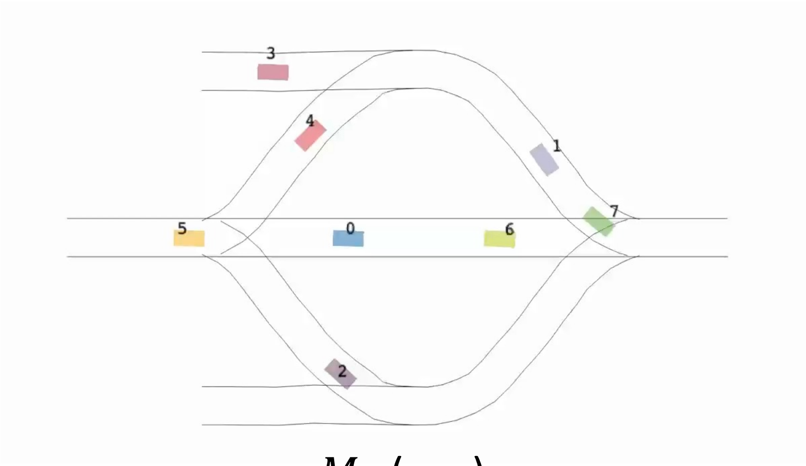


(c) With 8 agents

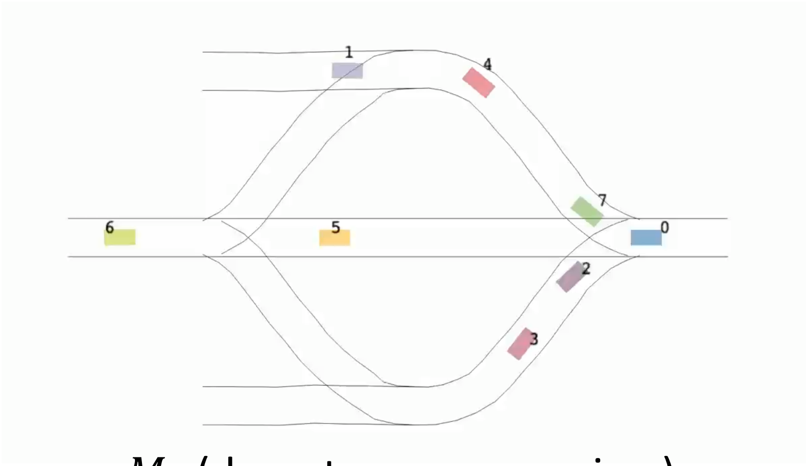


(d) With 8 agents

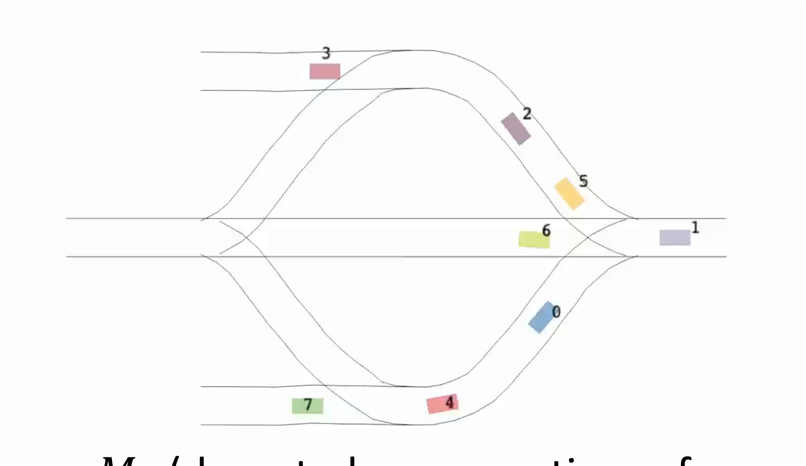
Evaluations | Testing Demo | Scenario (d)



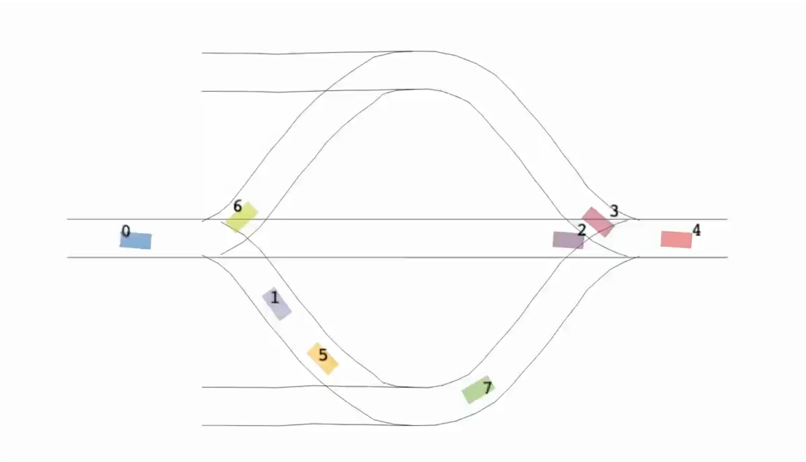
M_0 (our)



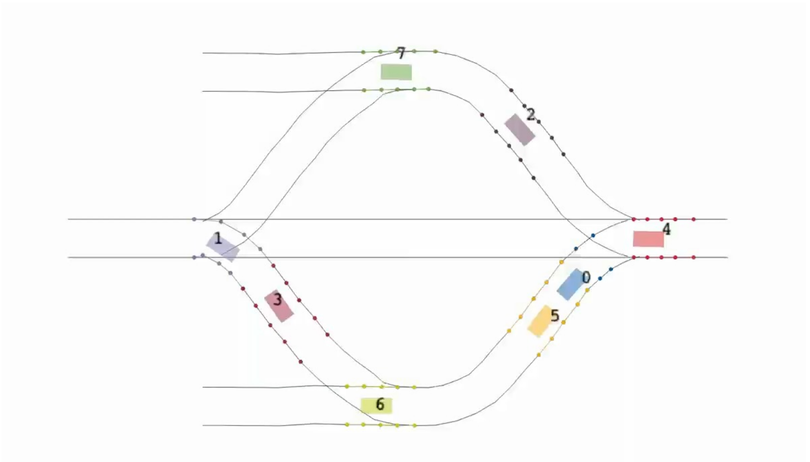
M_1 (do not use an ego view)



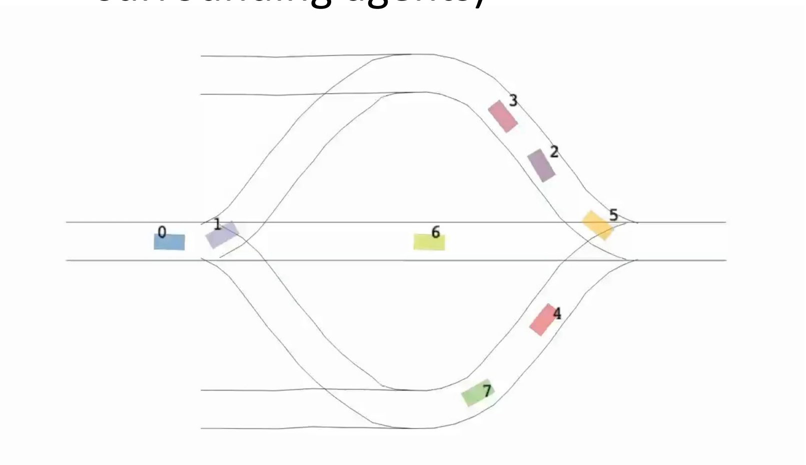
M_2 (do not observe vertices of surrounding agents)



M_3 (do not observe distances to surrounding agents)



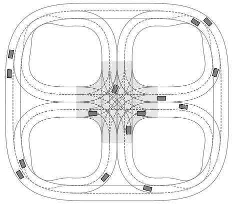
M_4 (do not observe distances to lane boundaries)

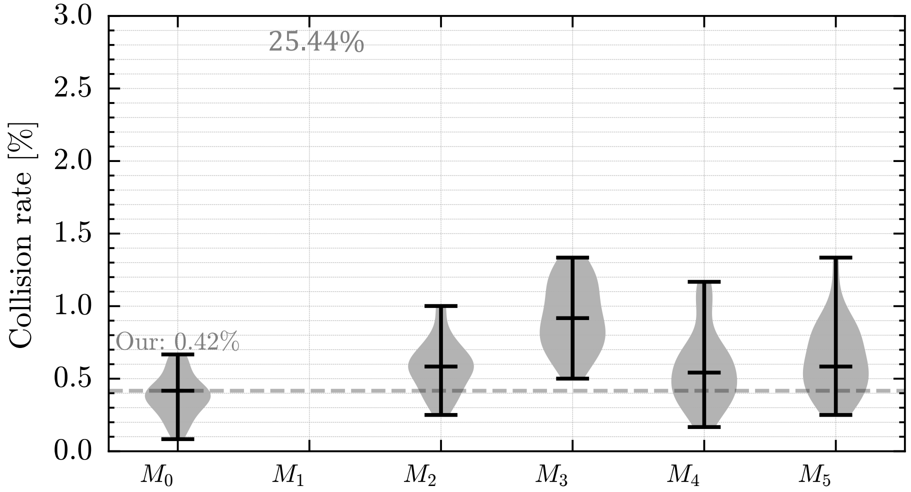


M_5 (do not observe distances to lane center lines)

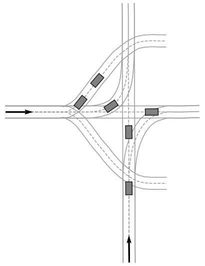
Evaluations | Collision Rate

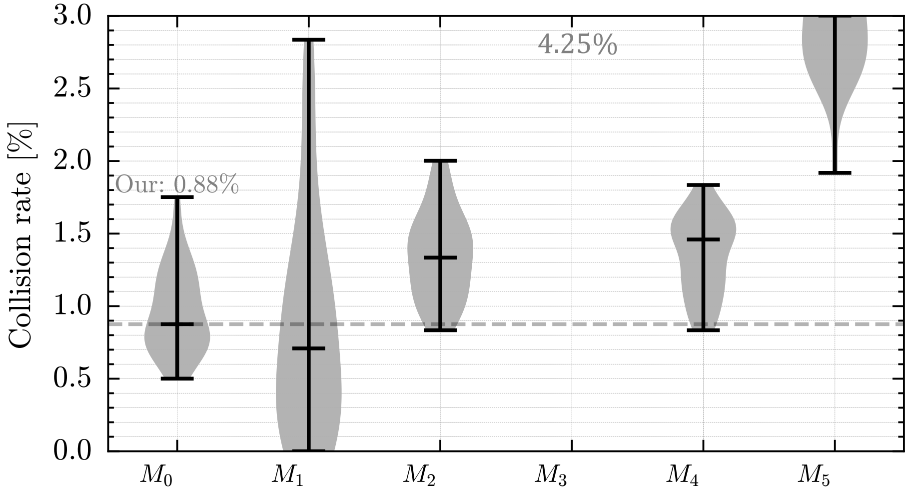
Scenario (a)



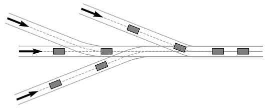


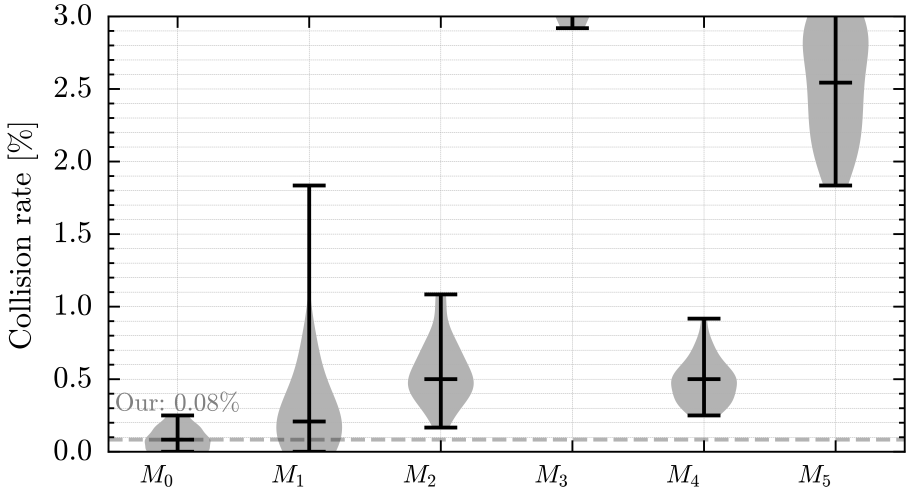
Scenario (b)



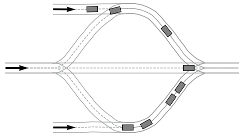


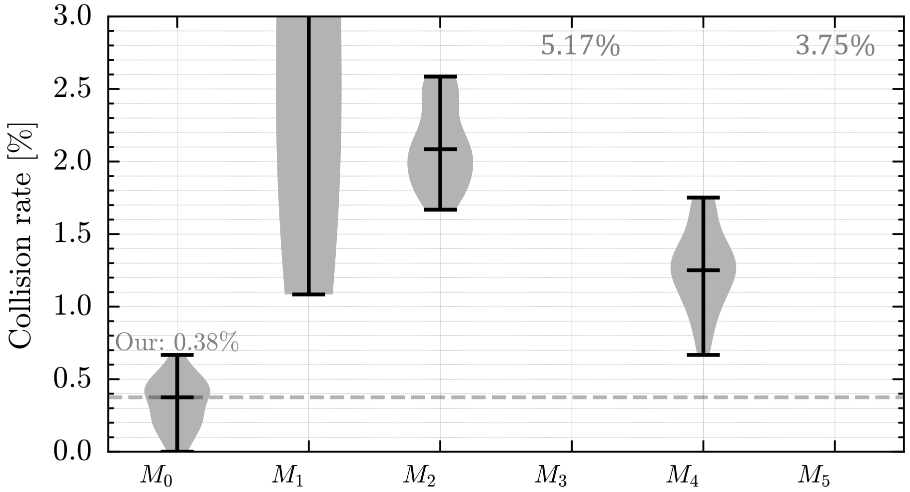
Scenario (c)





Scenario (d)



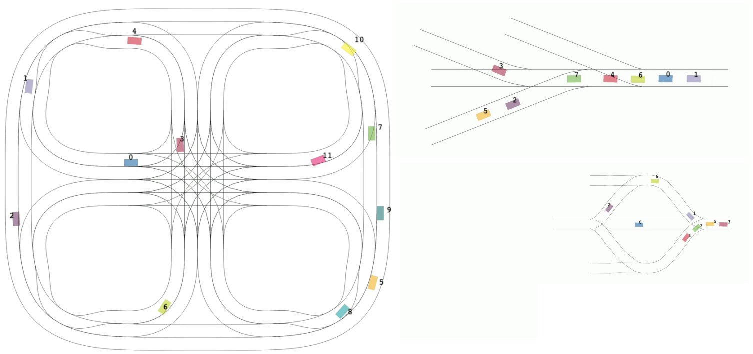


Summary

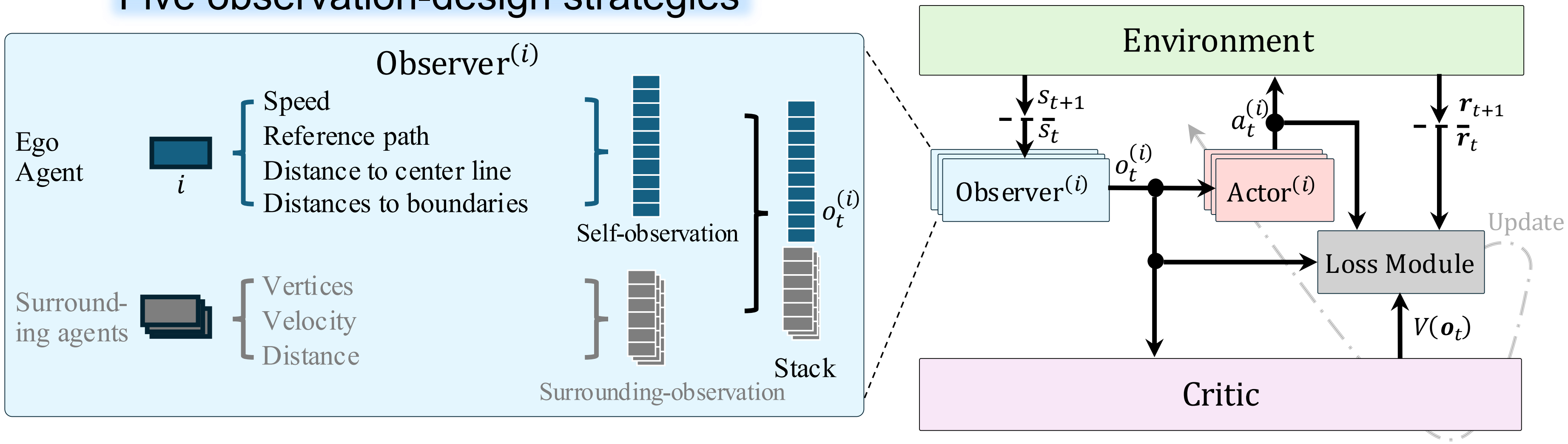
- **SigmaRL**, a sample-efficient and generalizable MARL Framework for motion planning of CAVs

<1 h training time on a single CPU

Completely unseen scenarios



Five observation-design strategies

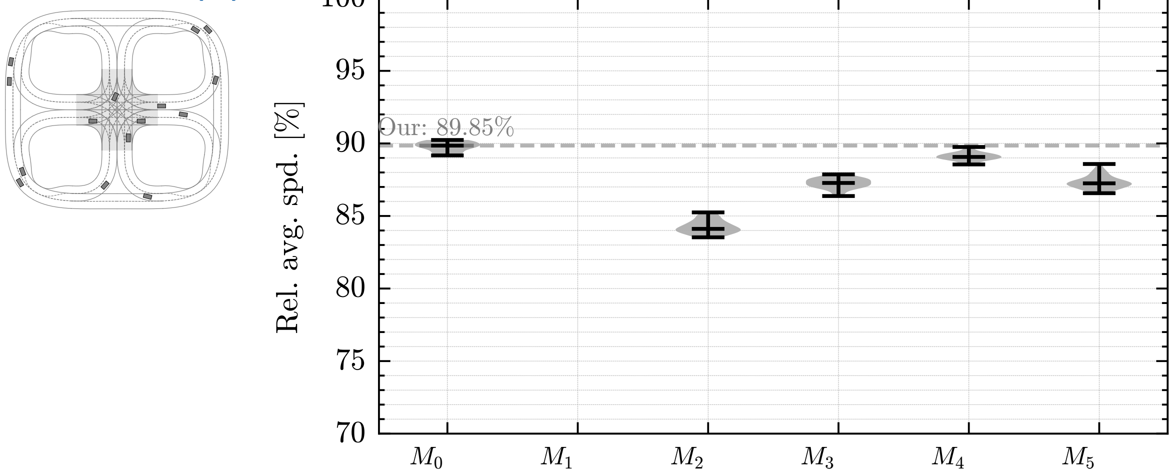


Preprint | Code | Video

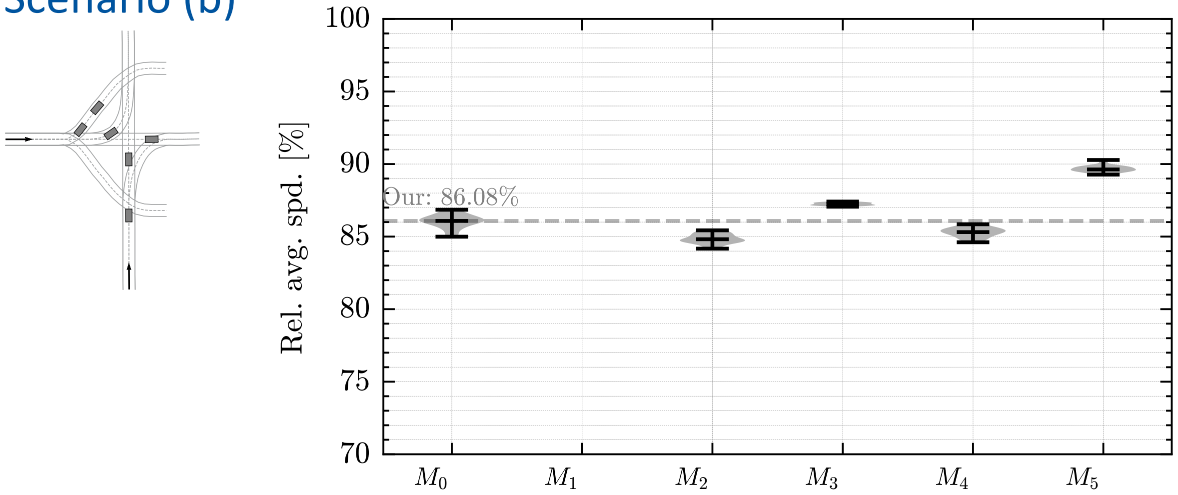
Backup

Evaluations | Average Speed

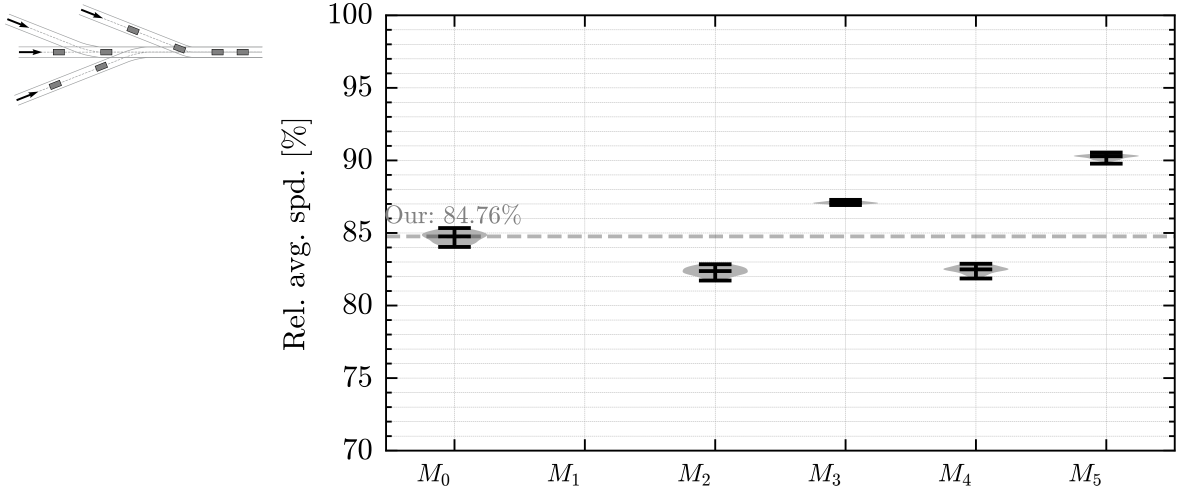
Scenario (a)



Scenario (b)



Scenario (c)



Scenario (d)