Learning Eco-Driving Strategies at Signalized Intersections

Massachusetts Institute of Technology

Vindula Jayawardana†, Cathy Wu‡

† EECS, LIDS   ‡ CEE, LIDS, IDSS

Abstract

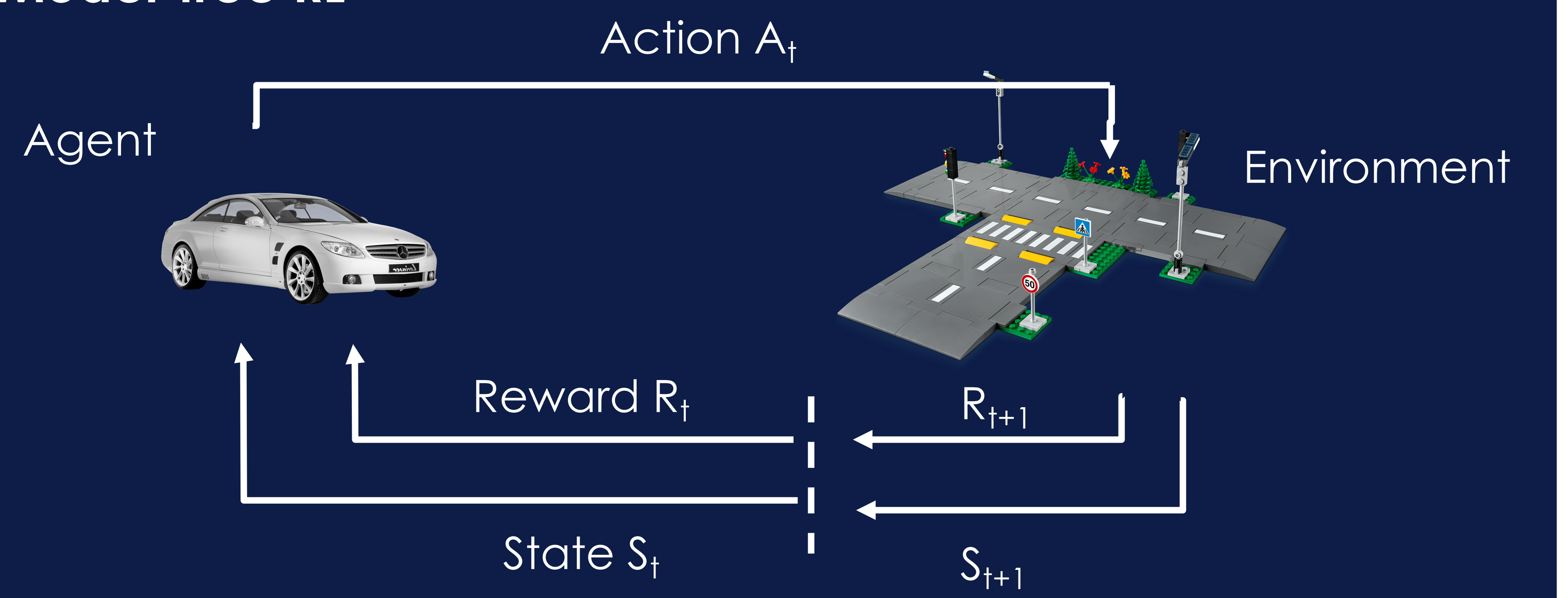
- Multi agent reinforcement learning approach to learn eco-driving strategies at signalized intersections.
- Under 100% penetration of CAVs,
  - 18% reduction in fuel
  - 25% reduction in CO_2 emission
  - 20% increase in travel speed
- Even 25% CAV penetration can bring at least 50% of the total fuel and emission reduction benefits.

Introduction

- Transportation sector in the US contributes **29%** to the green house gas emission (GHG) in which **77%** is due to land transportation.
- Previous studies on eco-driving at intersections,
  - assumes a model of the vehicle dynamics (model-based)
  - simplify the objective to fuel reduction and ignore travel time
  - Involve solving a non-linear optimization problem in real time
- Our reinforcement learning based approach is model-free and optimize fuel consumption while reducing impact on travel time.

Methodology

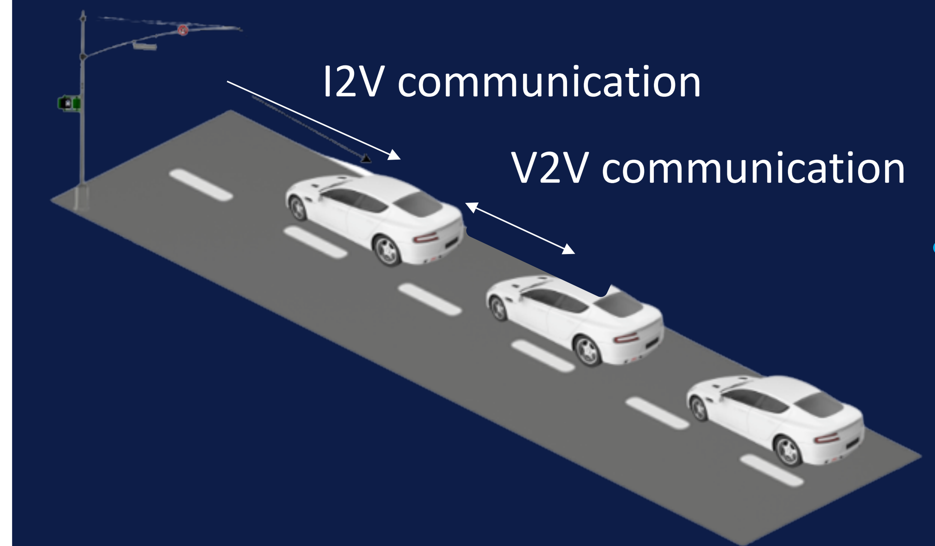
Model-free RL



Maximize discounted total reward $= \max_\theta \sum_{t=1}^{T} \gamma^t r_t(st, at = \pi_\theta(st))$

- In multi-agent RL, each agent has a policy

Eco-driving at signalized intersections



Markov Decision Process (MDP)

- State:
  - ego-vehicle velocity
  - ego-vehicle position
  - lead vehicle velocity
  - lead vehicle position
  - following vehicle velocity
  - following vehicle position
  - time to green
  - traffic phase
- Action:
  - longitudinal acceleration

  a ∈ (am_{ax}, ami_n)
- Reward:

$$r(s,a) = \begin{cases} R_1 & \text{if any vehicle stops at the start of a lane.} \\ R_2 & \text{if average fuel} \le \delta \wedge \text{average stops} = 0. \\ R_3 & \text{if average fuel} \le \delta \wedge \text{average stops} > 0 \\ R_4 & \text{otherwise} \end{cases}$$

- Challenges in composite reward design
  - objective terms are competing (fuel and travel time)
  - rate of change of the two reward terms are different in different regions of the composite objective

Results

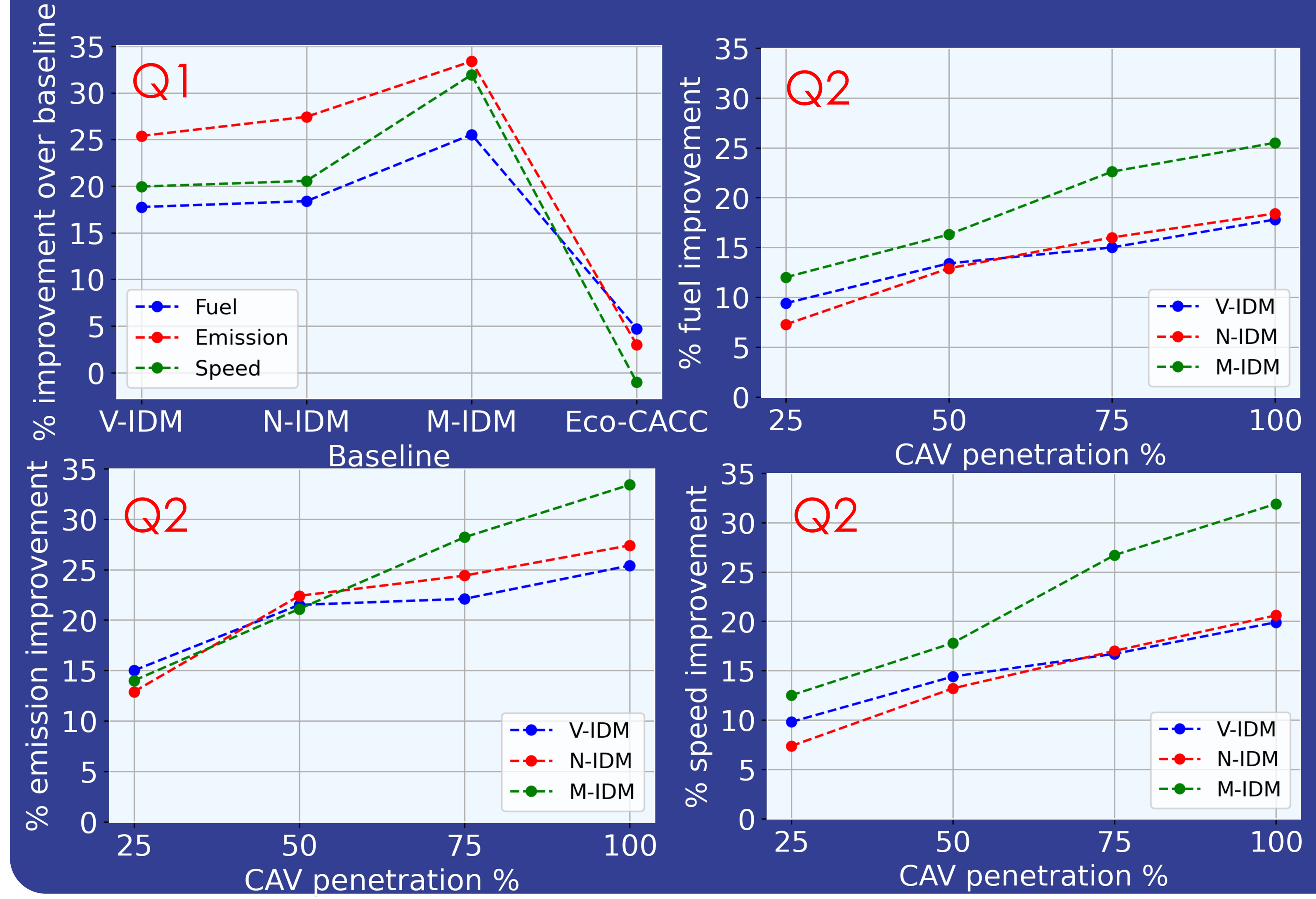
Baselines

- V-IDM: vanilla IDM car following model
- N-IDM: IDM model with noise (variability in driving)
- M-IDM: IDM with noise and varying parameters (diverse mix of drivers with varying levels of aggressiveness)
- Eco-CACC: a mode-based trajectory optimization

Fuel Model: VT-CPFM **Emission model:** HBEFA-v3.1

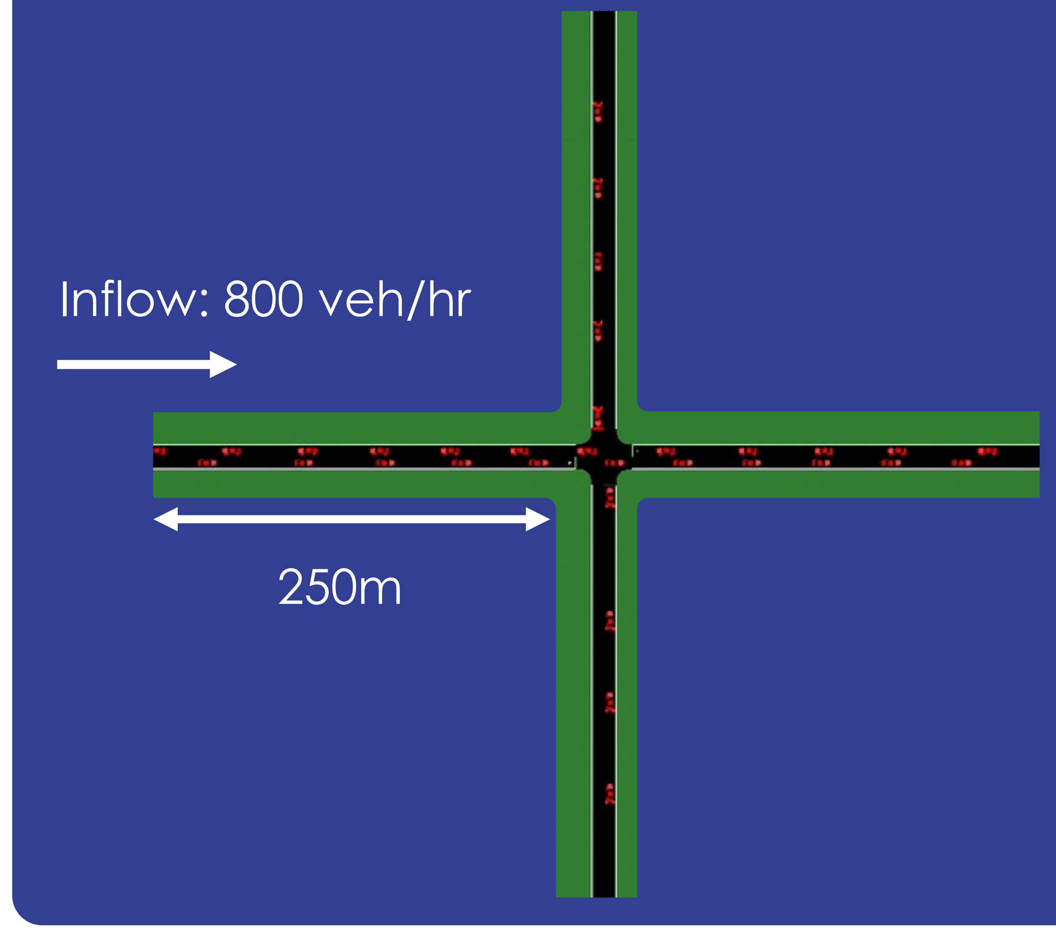
Questions

- Q1: How does the proposed control policy compare to naturalistic driving and model-based control baselines?
- Q2: How well does the proposed control policy generalize to environments unseen at training time?



Results

Learned behavior: 100% CAVs



Conclusion

- Significant savings in fuel, emission while even improving travel speed.
- Generalizability of learn policies to out- of-distribution settings is successful
- Future work: National level impact assessment as a climate change intervention

Acknowledgements

- MIT SuperCloud and Lincoln Laboratory Supercomputing Center.
- Mark Taylor, Blaine Leonard, Matt Luker and Michael Sheffield at the Utah Department of Transportation.