Learning to Trick Robots into Cooperative Behavior

Jen Jen Chung

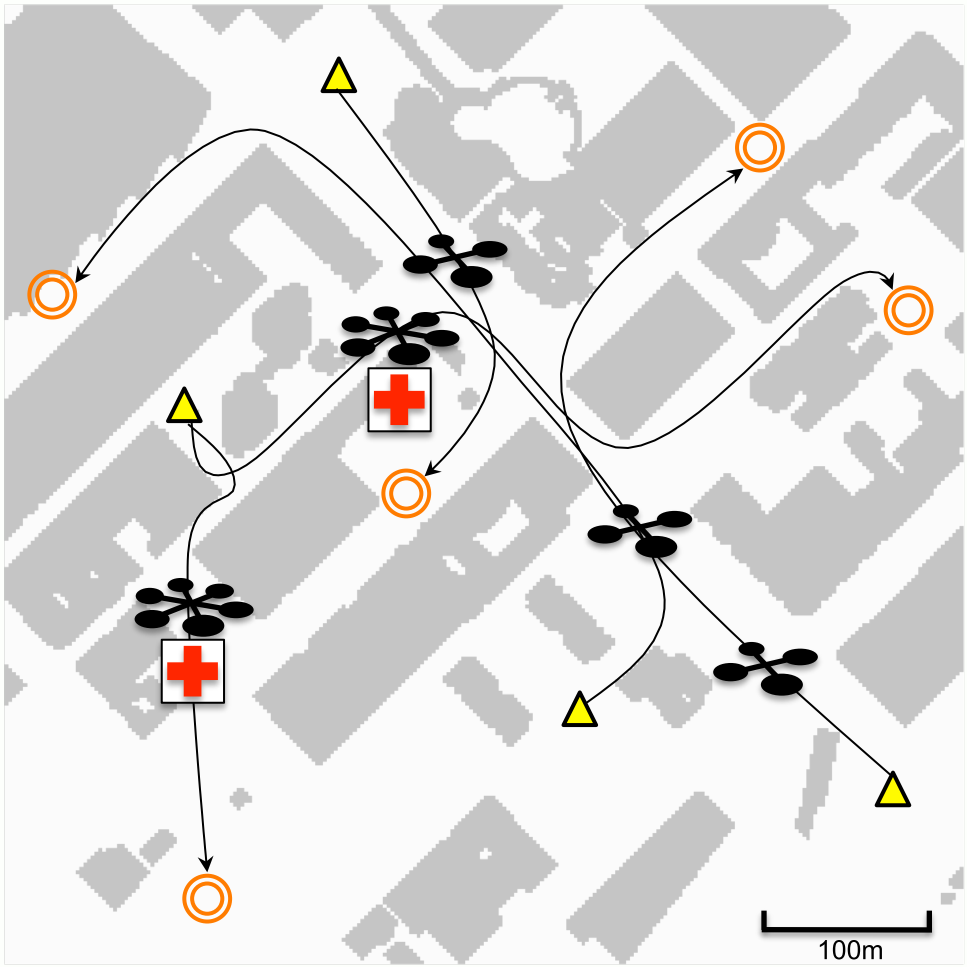
Autonomous Agents and Distributed Intelligence Lab

Oregon State University

UAV Package Delivery

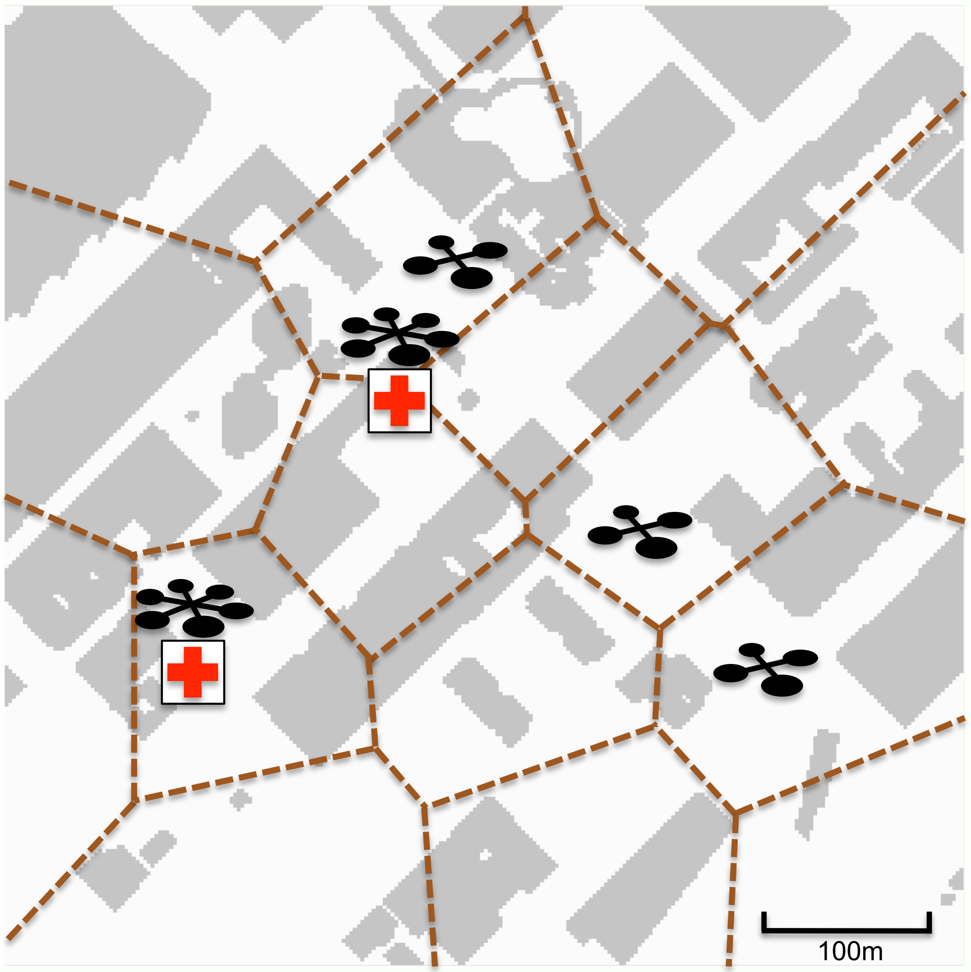
- Increasing interest in delivery drones: UPS, Amazon, etc.
- Dense UAV traffic in cluttered urban environment
- No current framework for large scale coordination

A Cross-Section of the Airspace



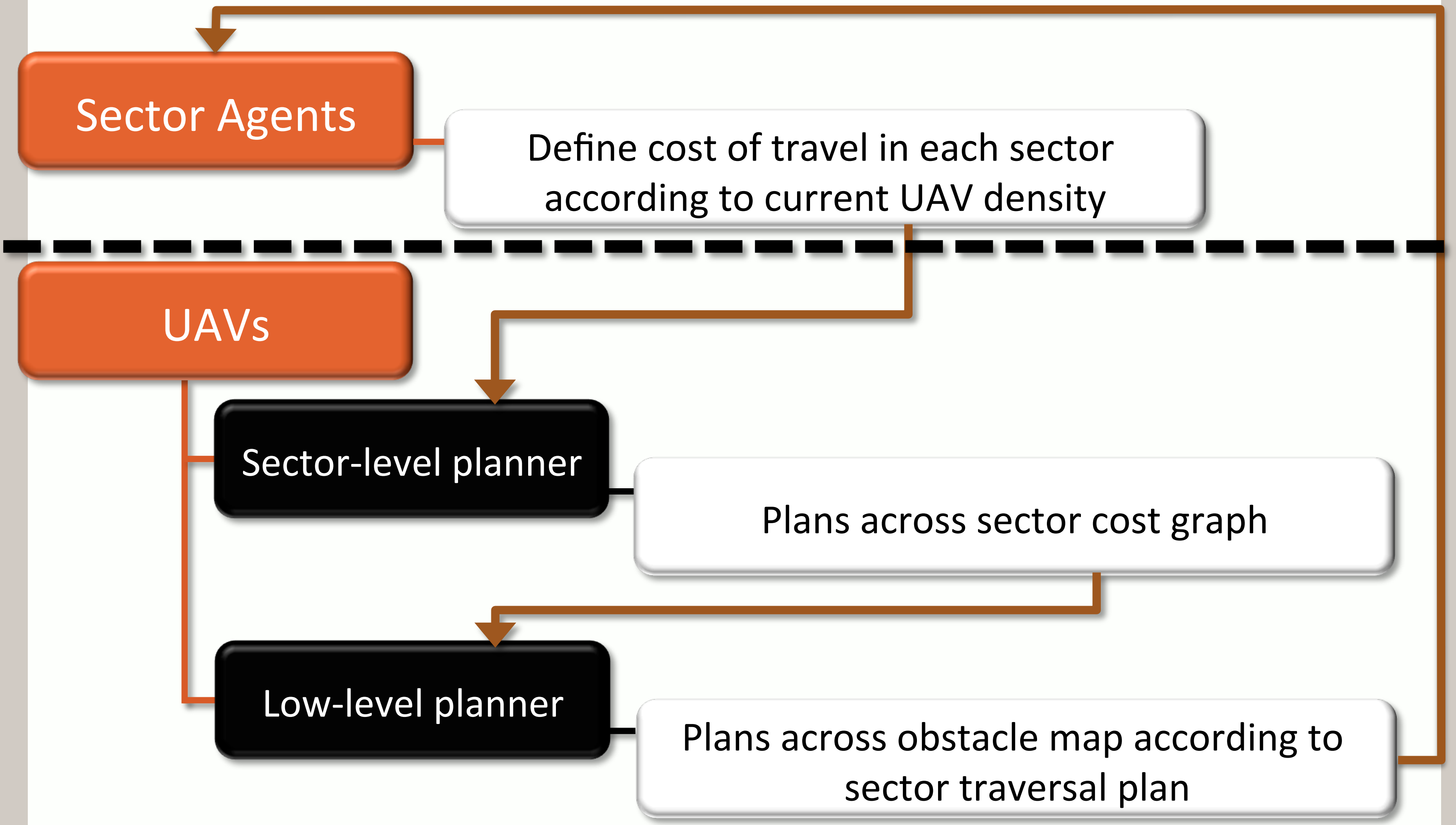
- Automated UAV traffic management
- Challenges:
  - Narrow thoroughfares of dense traffic
  - Heterogeneous UAVs
  - Dynamic obstacle landscape
- Goals
  - Minimize conflict occurrences
  - Avoid cascading effects
  - Maintain throughput

Multiagent UAV Traffic Management (UTM)



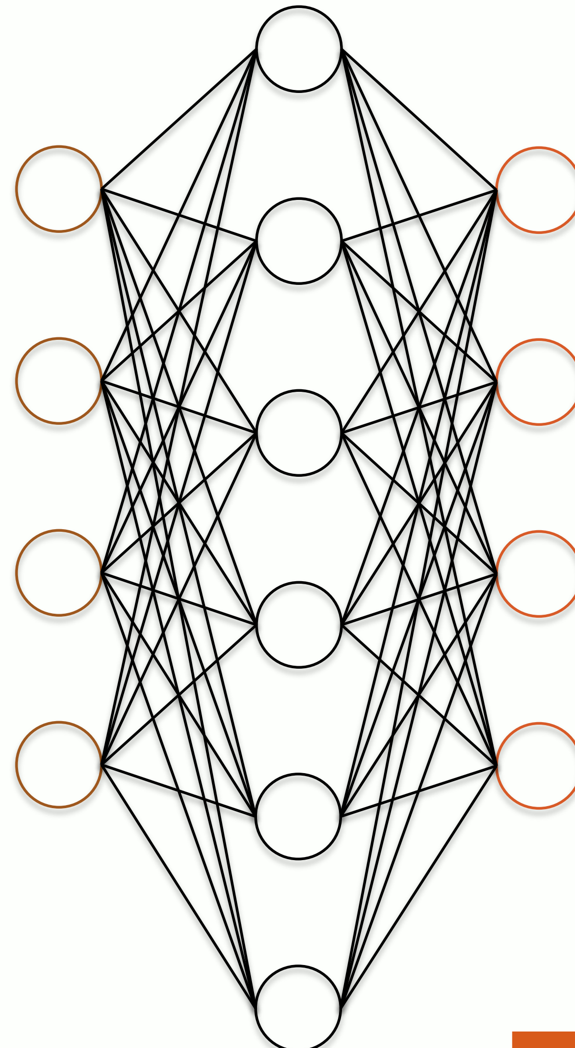
- Divide airspace into sectors
  - Assign single UTM agent to manage each sector
- Multiagent team:
  - UTM agents **individually** learn policy for assigning sector traversal costs
  - Reward is total number of conflicts in **global** system

A Hierarchical Approach

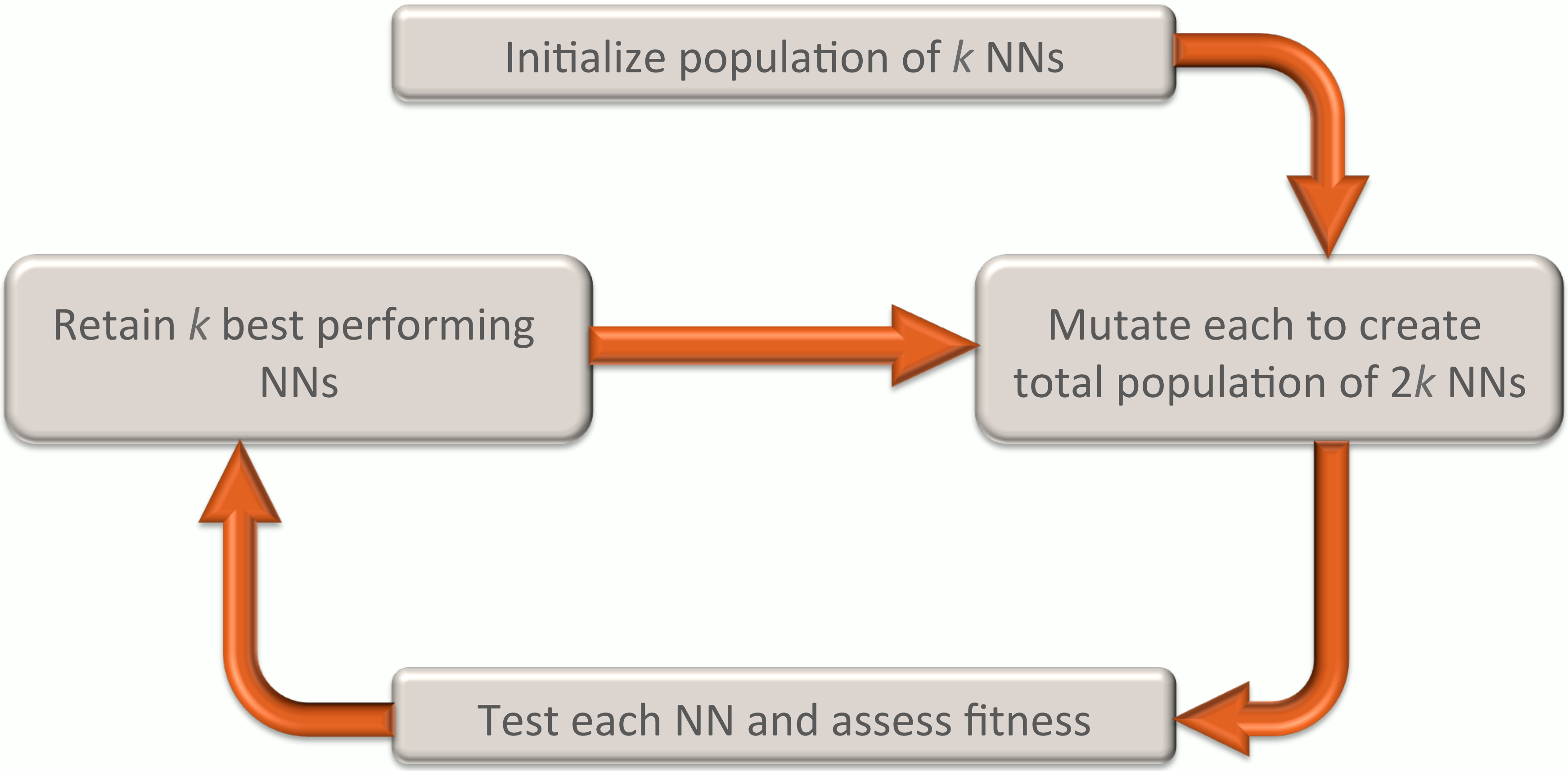


UTM Learning Agents

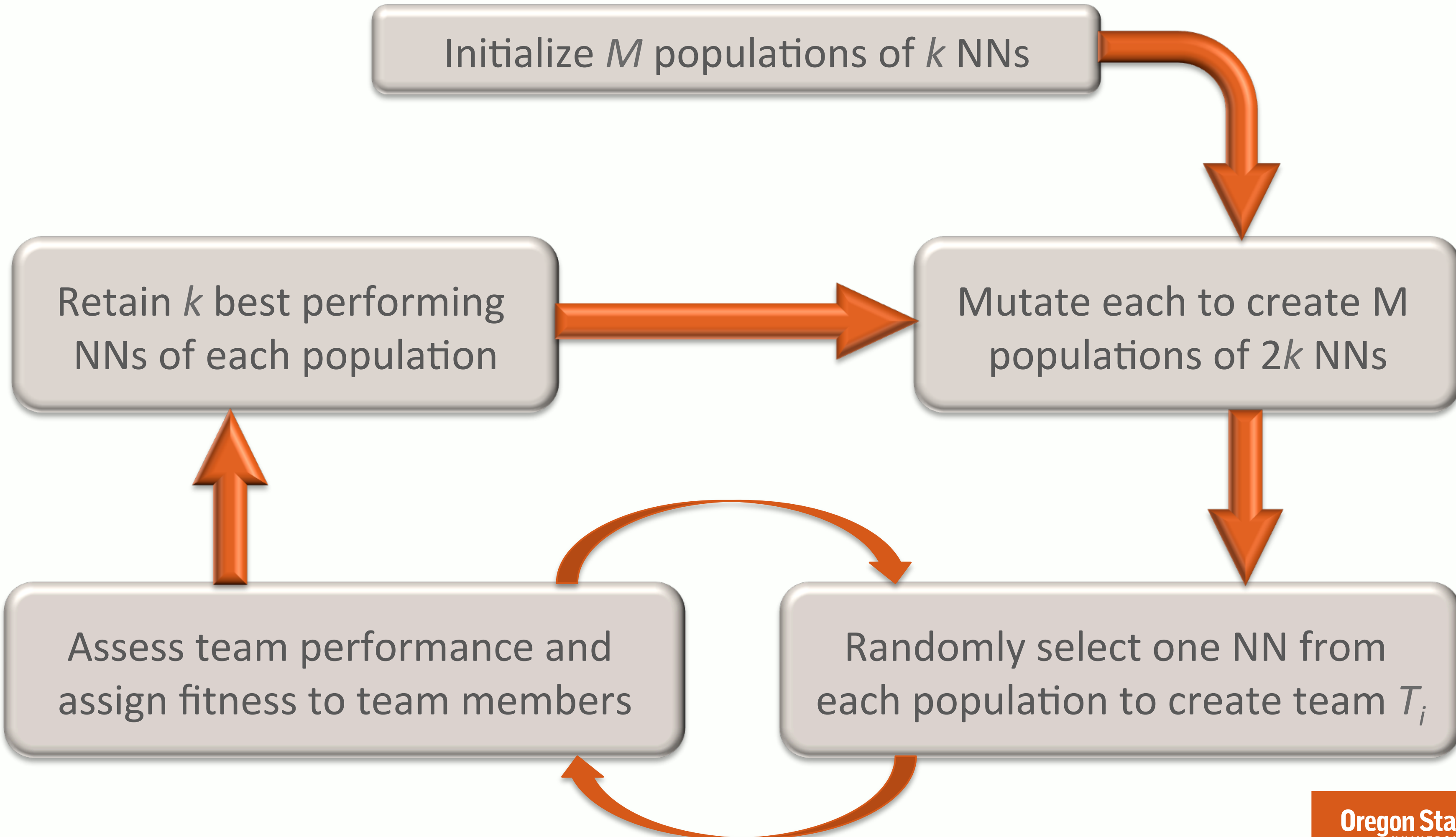
- Learn the cost of travel to apply to UAVs in the sector
- Neural network control
  - Inputs: UAV counts in sector
    - Separate into traffic types, e.g. heading, priority, platform etc.
  - Outputs: Cost of through-sector travel for each traffic type
- Cooperative coevolution to learn NN weights
  - Fitness value: number of conflicts



Evolutionary Algorithms for Learning Control Policies



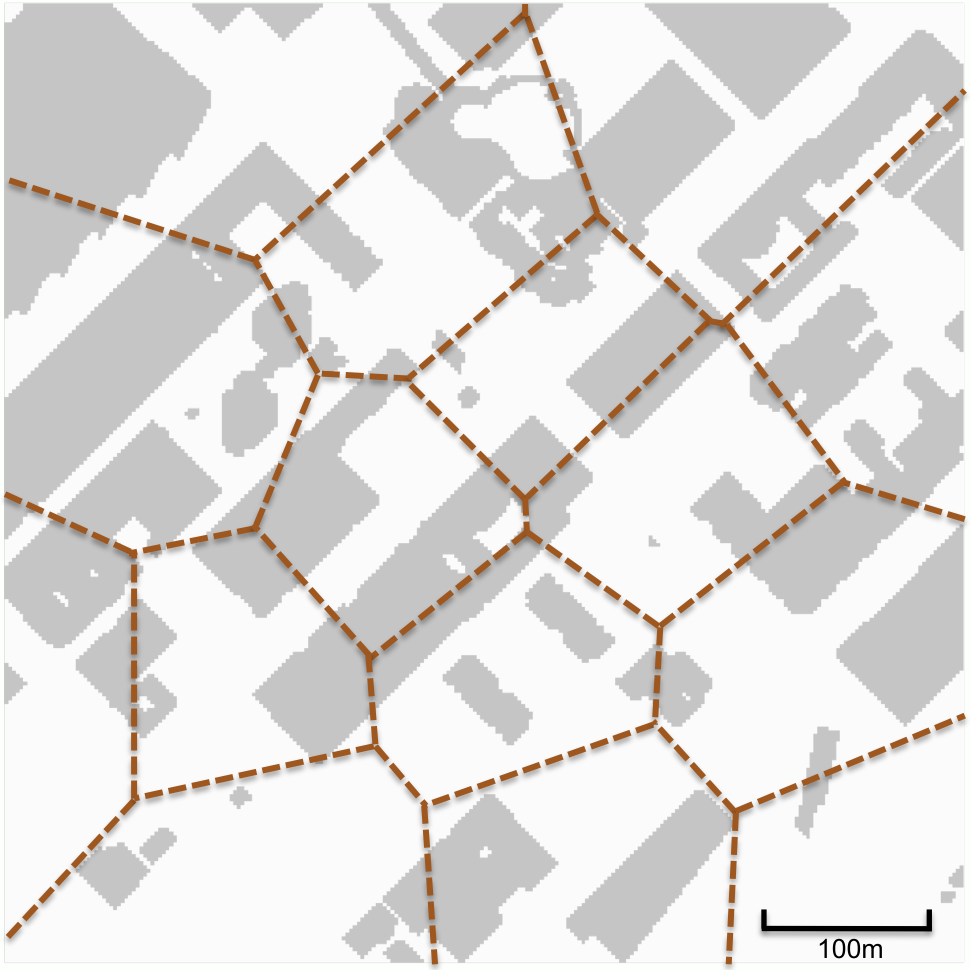
Cooperative Coevolutionary Algorithms (CCEAs)



Simulation Experiments

- Urban airspace
  - 256×256 cell map of San Francisco
  - 15 Voronoi partitions

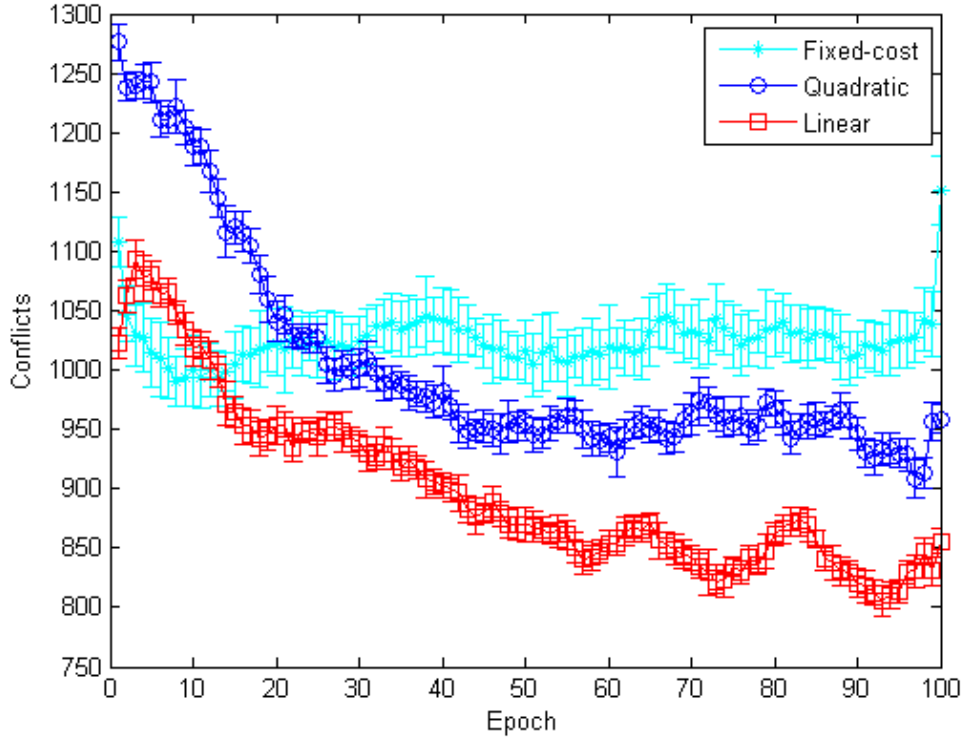
- Fitness calculation
  - Linear: no. conflicts at each cell summed
  - Quadratic: no. conflicts at each cell squared and summed



Simulation Experiments

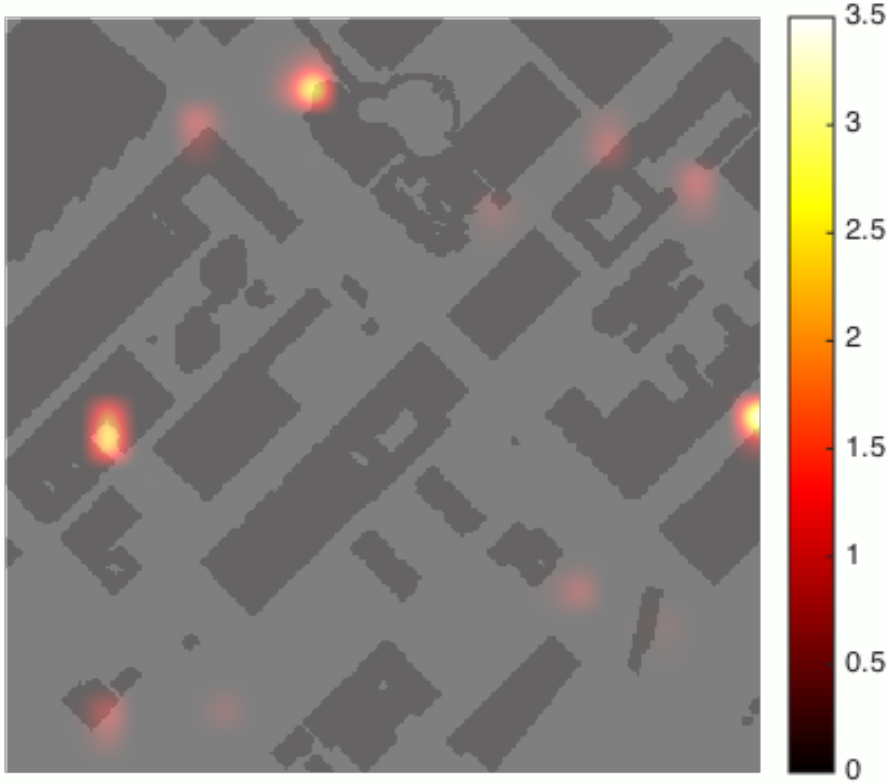
- Sector agents
  - Initialized with population of 10 NN control policies, 10% mutation noise
  - Inputs: $\{n_N, n_S, n_E, n_W\}$
  - Outputs: $\{c_N, c_S, c_E, c_W\}$
  - Fitness: number of conflicts
- UAVs
  - Stochastically generated from predefined set of start and goal locations
  - Approximately 100 UAVs in airspace during single learning epoch
  - A* planning at both sector- and low-level
  - Conflict radius: 2 cells (approx. 4m)

Learning Results: Total Conflicts

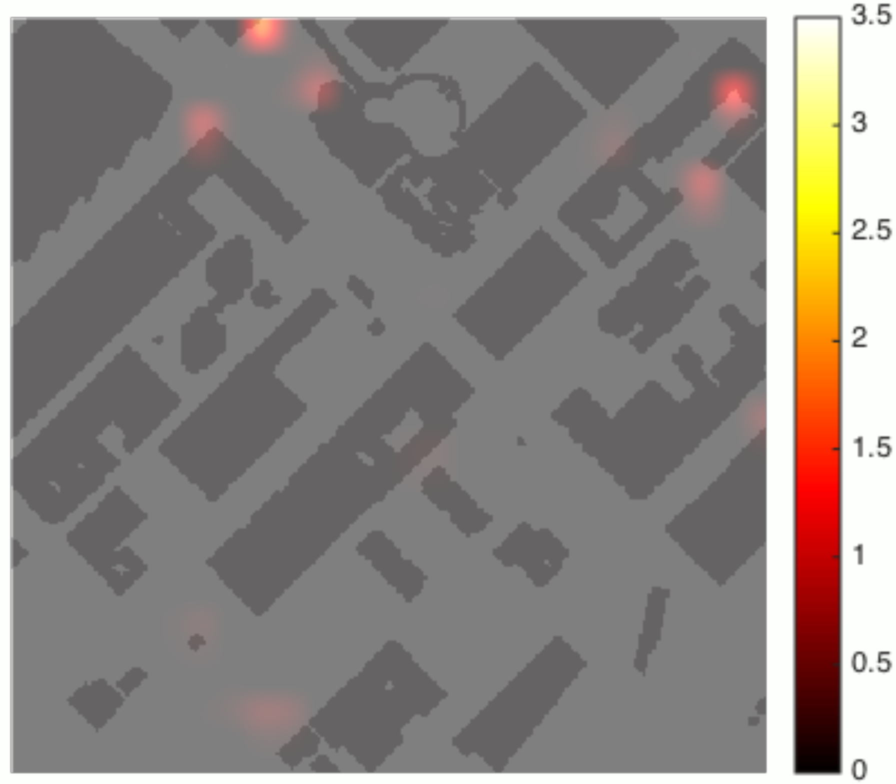


- Team performance over 100 learning epochs
- Averaged over 20 trials
- 16% reduction in total system conflicts

Congestion Reduction: Linear Cost Fitness Function



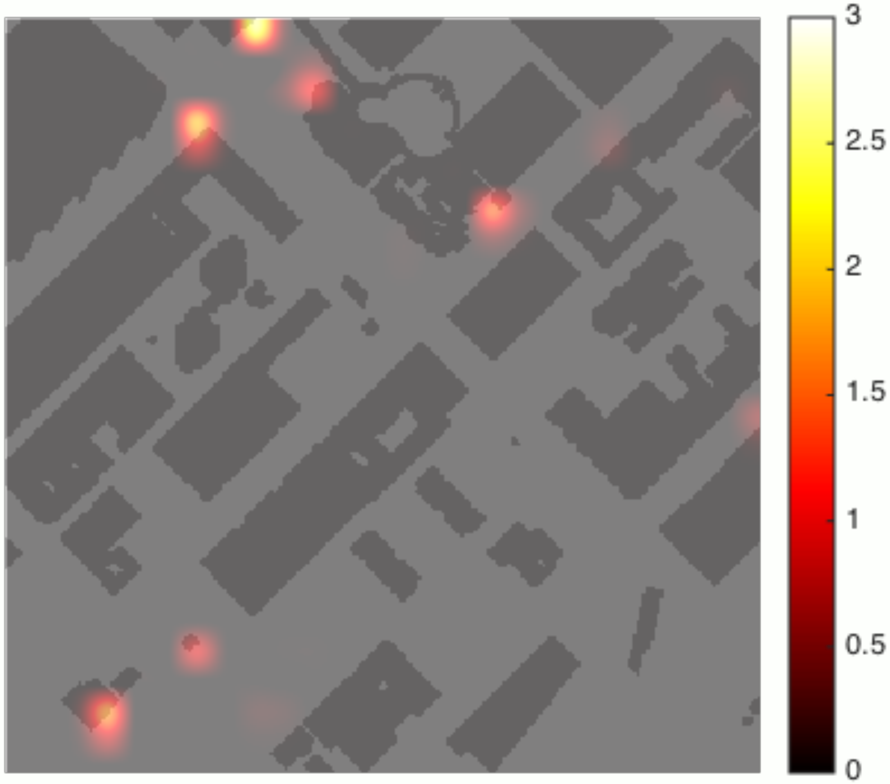
Random initialized sector costs



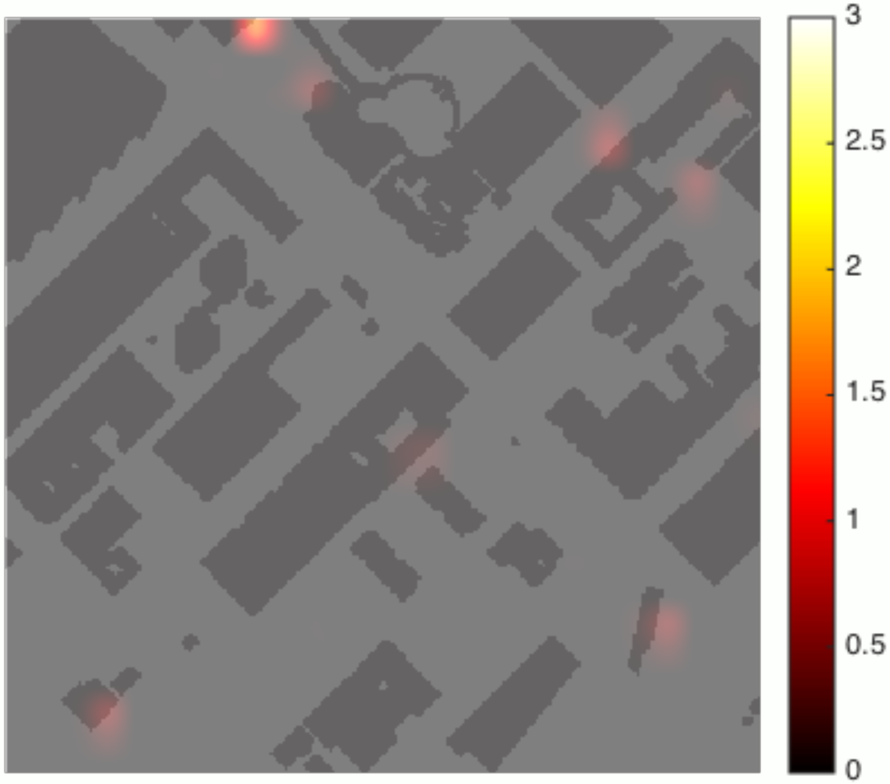
Learned sector costs

Oregon State UNIVERSITY

Congestion Reduction: Quadratic Cost Fitness Function



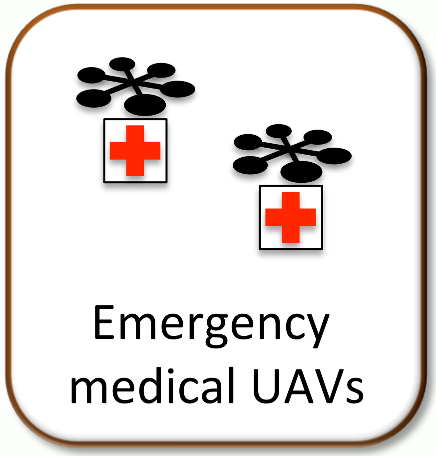
Random initialized sector costs



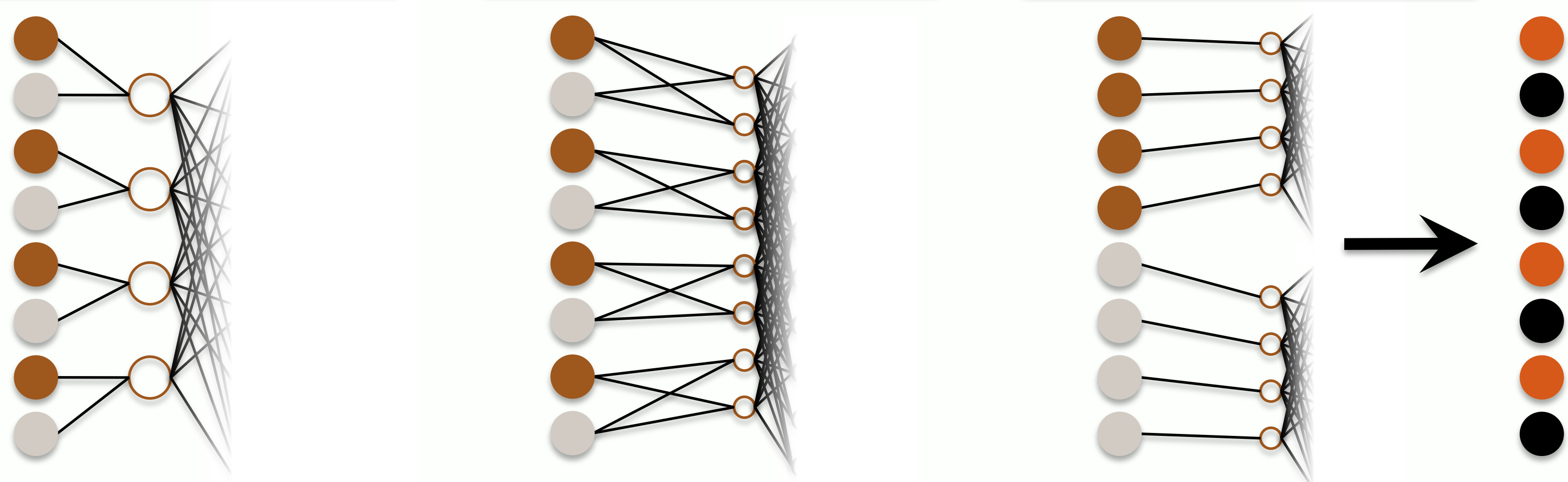
Learned sector costs

Extensions to Sector Agent Control Policies

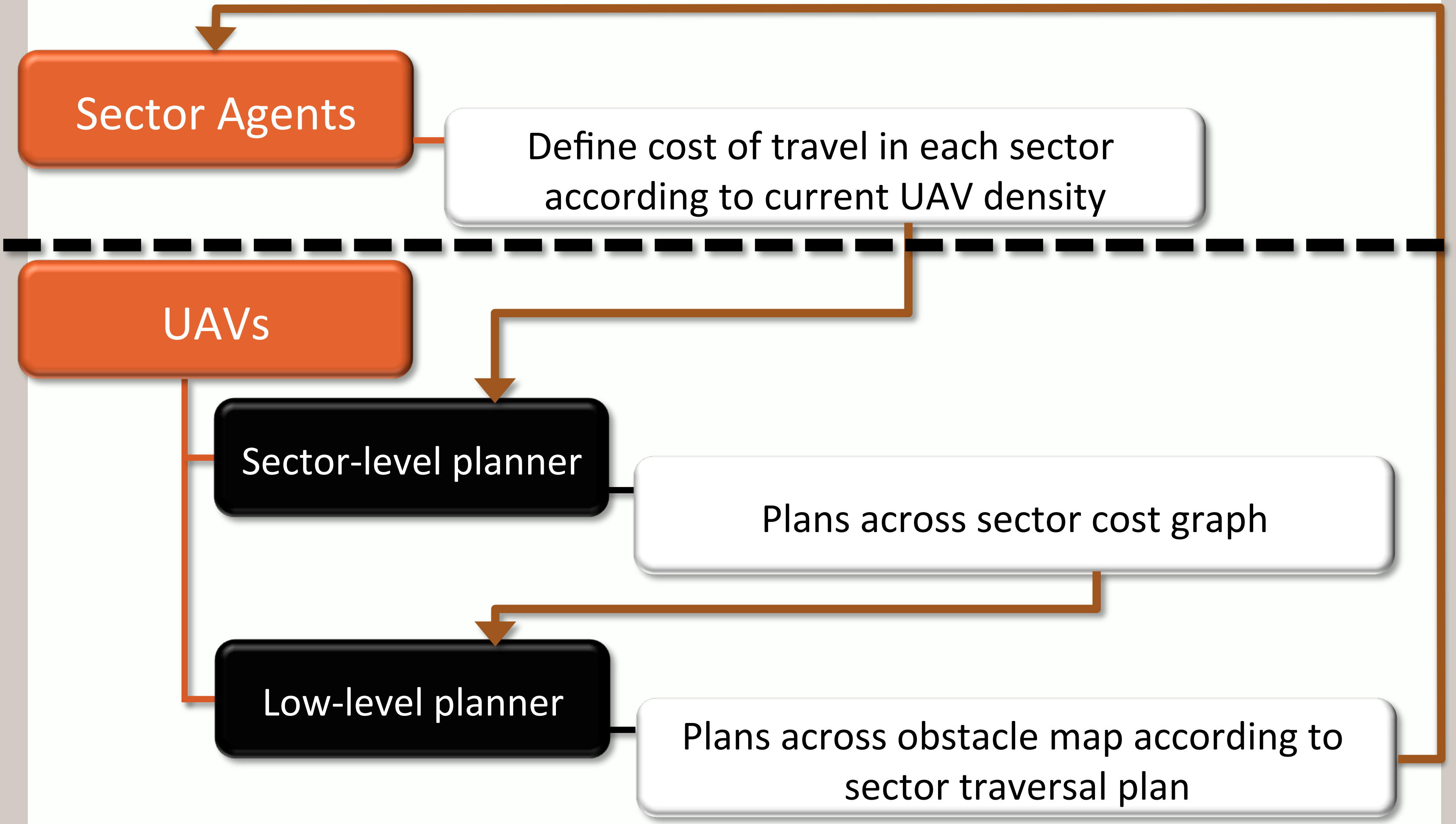
- Not all UAVs in the airspace are equal
- Account for UAV type in NN inputs and outputs



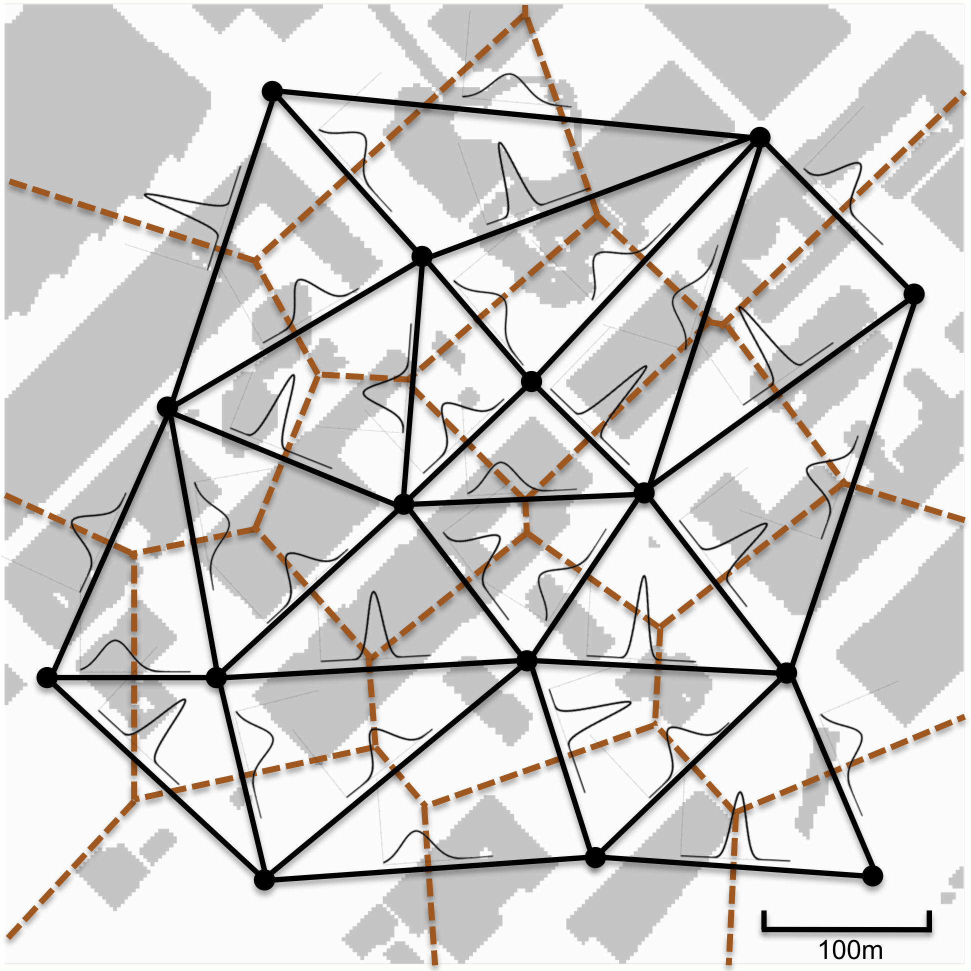
Weighted | Cross-weighted | Multi-mind



A Hierarchical Approach



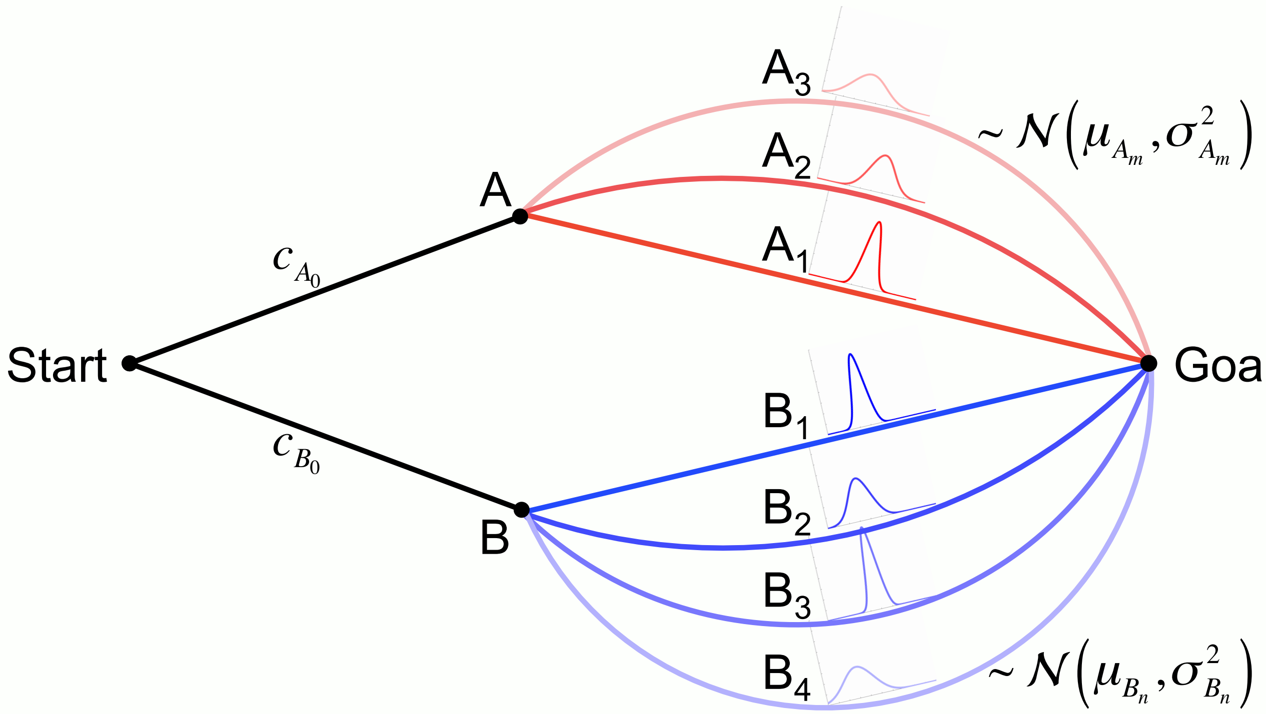
Risk-Aware Graph Search (RAGS)



- Graph search with uncertain edge costs
  - Normal distributions
- Bound path set
  - Domination according to mean and variance

$$A < B \leftrightarrow \left(A.c < B.c\right) \wedge \left(A.\sigma^2 < B.\sigma^2\right)$$

RAGS Path Execution



The probability that traveling via B
will yield a cheaper path than traveling via A

$$\int_{-\infty}^{\infty} \sum_{i=1}^{m} P\left(c_{A_i} = x; c_{A_j} > x, \forall j \neq i\right) \cdot 1 - P\left(c_{B_i} > x, \forall i \in \{1, \cdots, n\}\right) dx$$

RAGS vs. Existing Planning Algorithms

- Testing on graph with 100 vertices
  - 3 sets of edge cost distributions

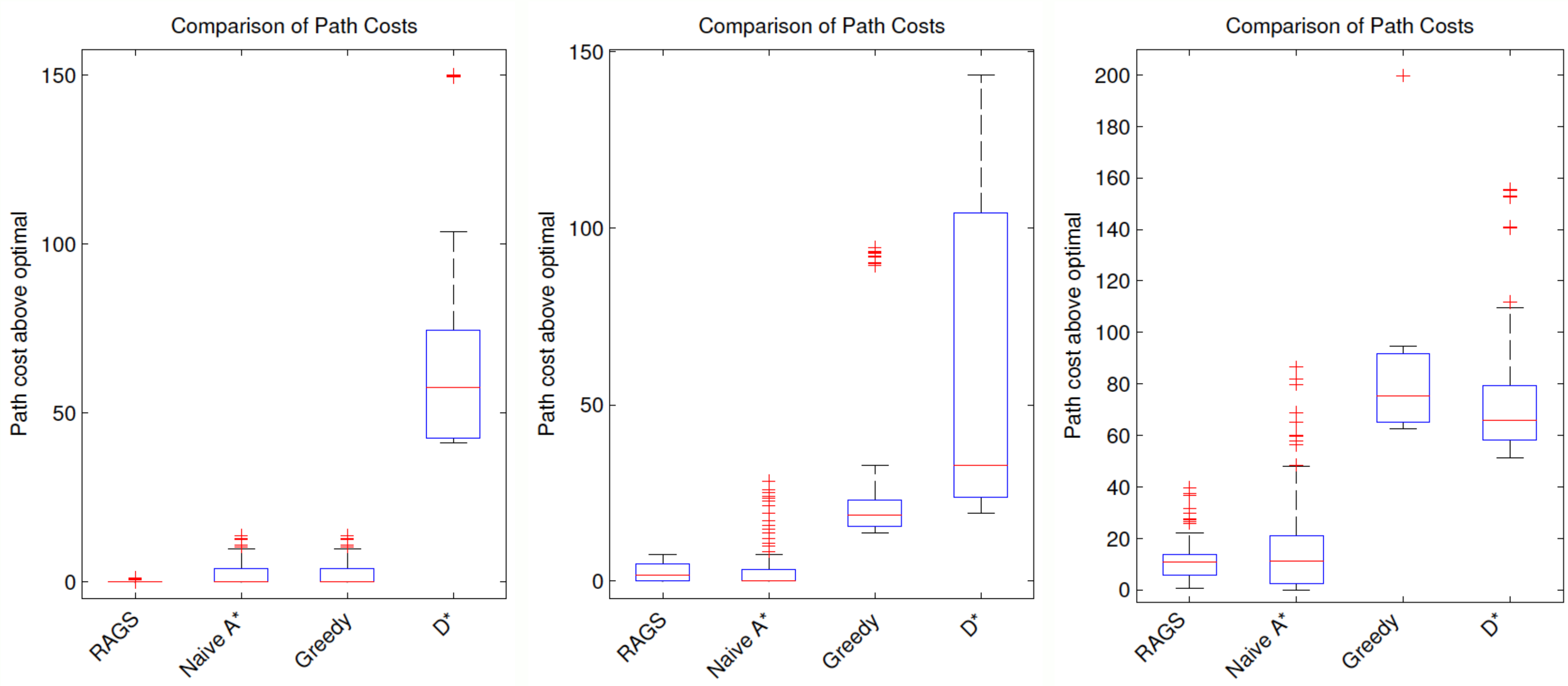
$$\text{Edge cost} = \text{Euclidean distance} + \varepsilon, \quad \varepsilon \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$\mu \in [0, 100]$$

$$\sigma^2 \in \left[0, \sigma_{\max}^2\right], \quad \sigma_{\max}^2 = \{5, 10, 20\}$$

- Compared against
  - Naïve A* on the mean
  - Greedy on bounded path set
  - D*

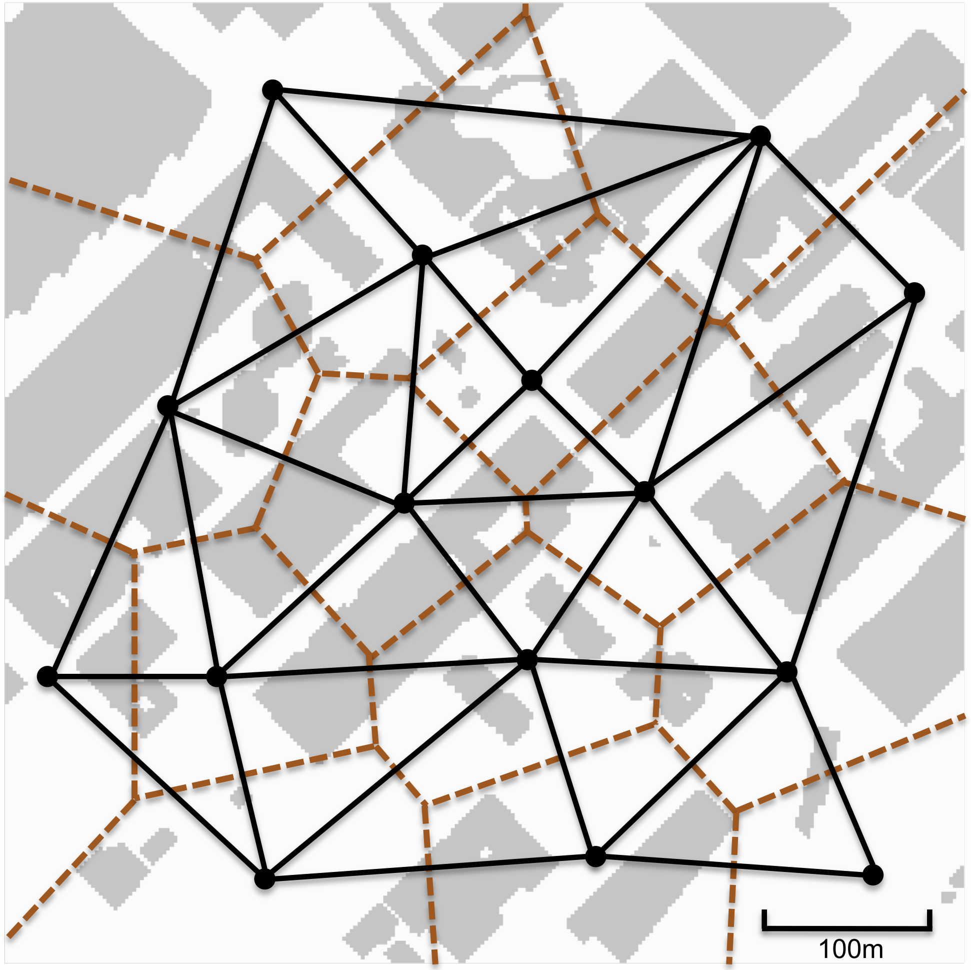
RAGS vs. Existing Planning Algorithms



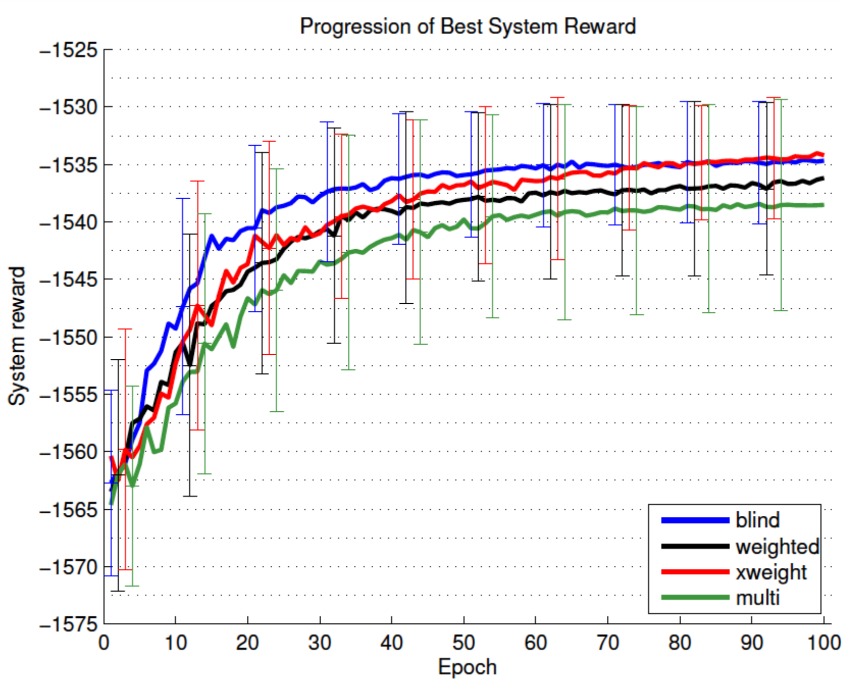
$\sigma^2 \in (0,5)$ $\sigma^2 \in (0,10)$ $\sigma^2 \in (0,20)$

Oregon State UNIVERSITY

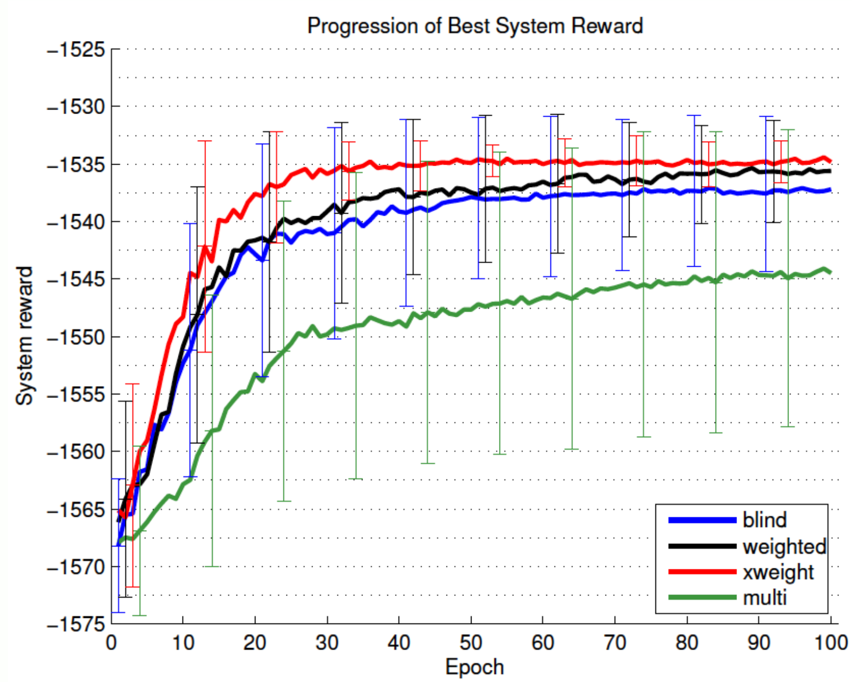
RAGS Integration with UTM Agents



Comparison of A* and RAGS



UAVs planning with A*



UAVs planning with RAGS

Conclusions and Future Work

- Implicit cooperation by learning individual control policies trained on global reward structures
- Risk-aware graph search accounts for modeled uncertainties in the environment
- Initial integration of high and low-level decision making shows faster learning rates
- Future work
  - Reward shaping to improve UTM agent policies
  - Theoretical guarantees of RAGS
  - Validation and verification

Acknowledgements

Professors

Graduate Students

Undergrads

Interns