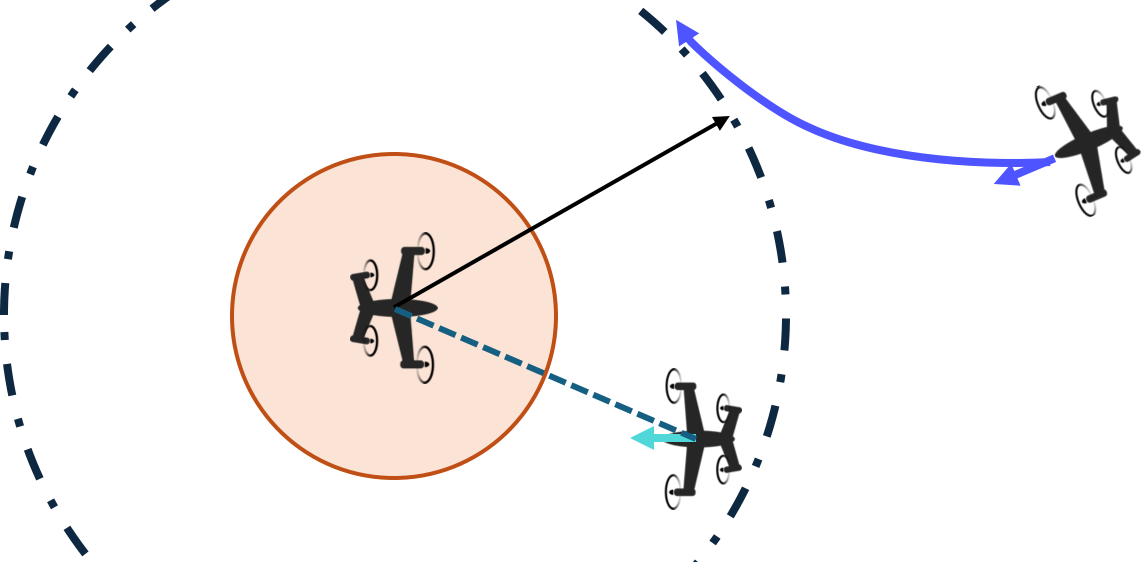


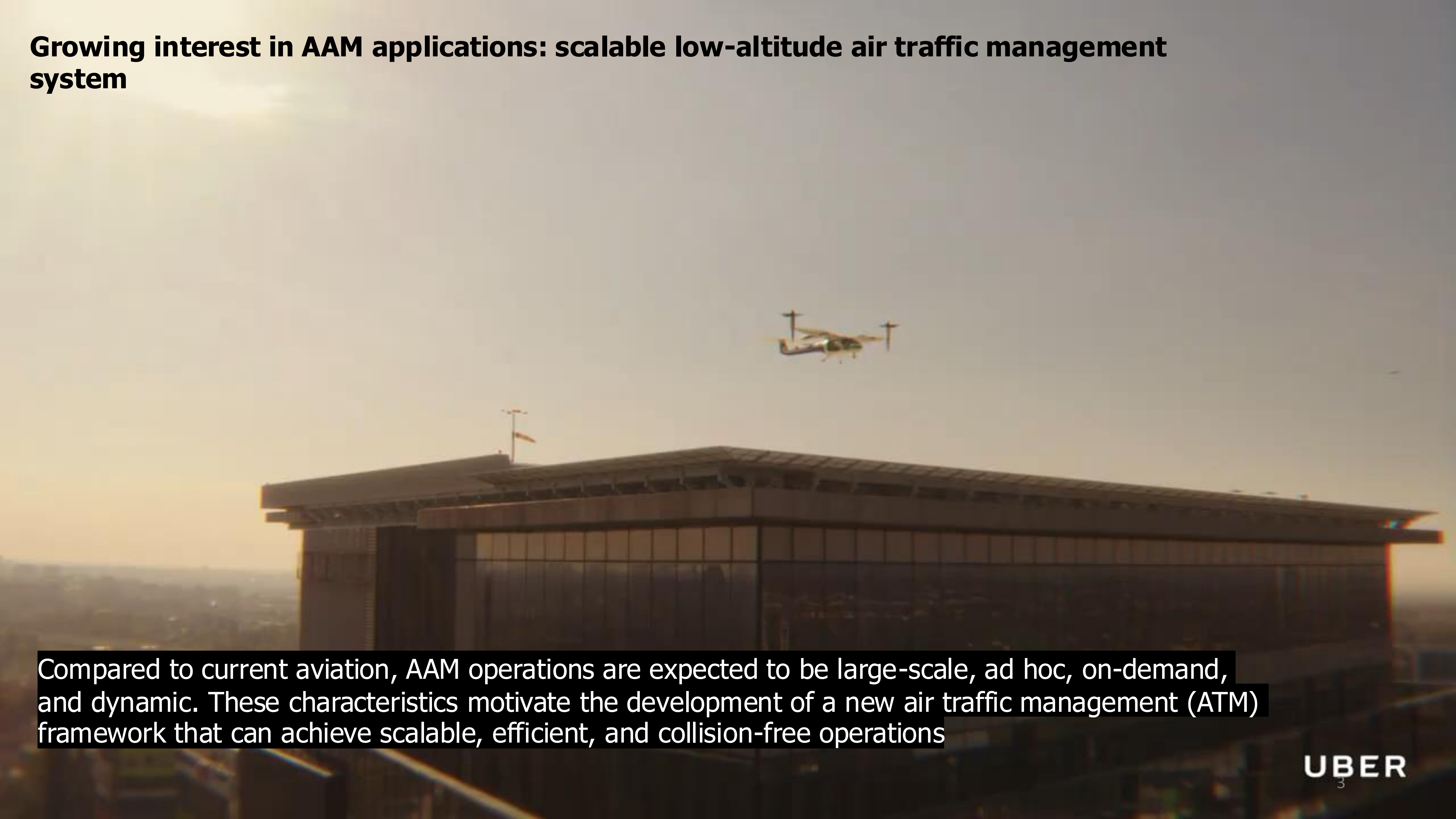
# Resolving Conflicting Constraints in Multi-Agent Reinforcement Learning with Layered Safety

Jason J. Choi*, Jasmine Jerry Aloor*, Jingqi Li*, Maria G. Mendoza,
Hamsa Balakrishnan, Claire J. Tomlin

# Collision-free operation is a fundamental requirement

for multi-robot coordination tasks
formation control,
multi-robot payload transport,
and autonomous navigation.

# Growing interest in AAM applications: scalable low-altitude air traffic management system



Compared to current aviation, AAM operations are expected to be large-scale, ad hoc, on-demand, and dynamic. These characteristics motivate the development of a new air traffic management (ATM) framework that can achieve scalable, efficient, and collision-free operations

A lot of safety measures are being practiced in the current aviation industry to maintain safety.

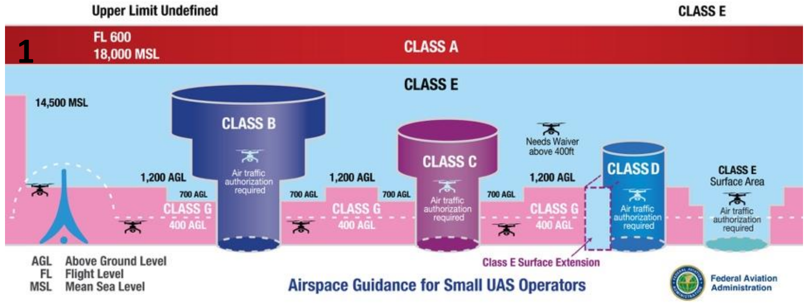
# Previous work:
# **Deconfliction & Collision Avoidance for Aviation**

Airspace Separation[1]

Centralized manual Air Traffic Control (ATC)[2]

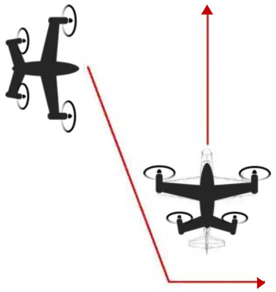
Protocols & Conventional Practice[3] built around assumptions that ***at most two vehicles*** engage in deconfliction

Imminent collision avoidance autopilot (TCAS, ACAS[4])
(20~30 sec before collision)





3





1 FAA, Airspace 101 – Rules of the Sky 3 Right-of-way rules, FAR 91.113 4 Kochenderfer et al., Lincoln Lab, 2012
Image source: FAA, SamChui, CFI Notebook

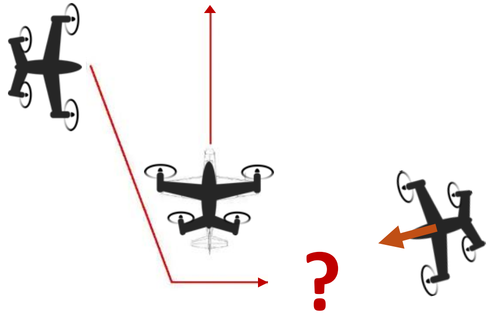
# Era of **Advanced Aerial Mobility:** New requirements

Airspace Separation

**Decentralized Air Traffic Operation**[1]

Protocols & Conventional Practice built around assumptions that ~~***at most two vehicles*** engage in deconfliction~~

Imminent collision avoidance autopilot (TCAS, ACAS)
(20~30 sec before collision)






1 NASA Ames, Provider of Services to UAM, 2021

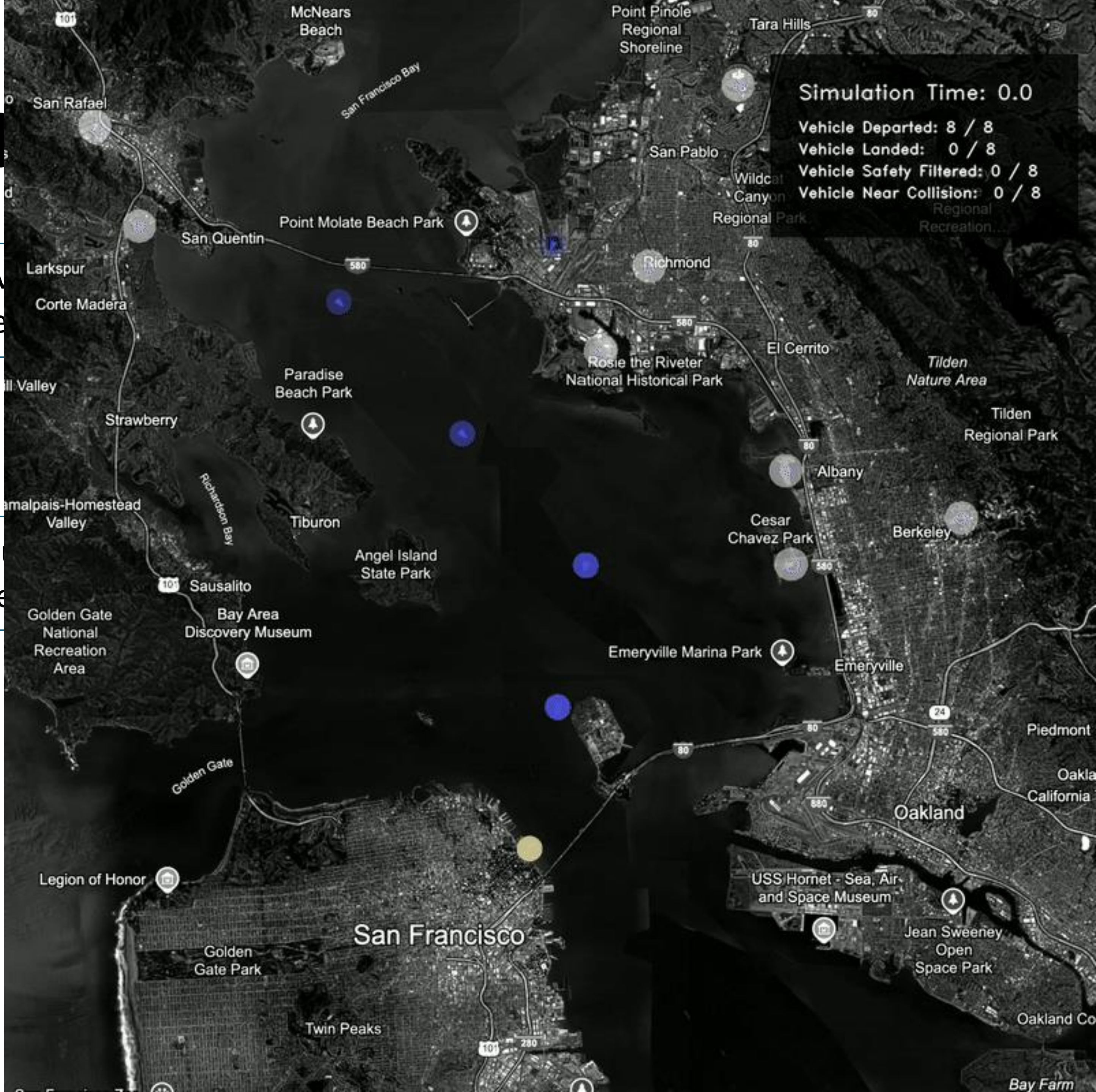
Image source: FAA

# Contributions: Layered

**1. Layered Safety Architecture:** Integration of CBV
reinforcement learning, creating a two-layer syste

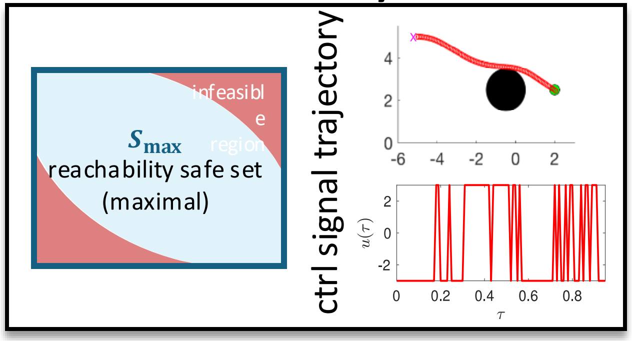
**2. Safe MARL Training Method:** Train using curric
conflicting constraint zones, without relying on pe

**3. Real-World & Simulated Validation:** Demonstrate effectiveness through hardware experiments with Crazyflie drones and simulations in dense AAM scenarios.

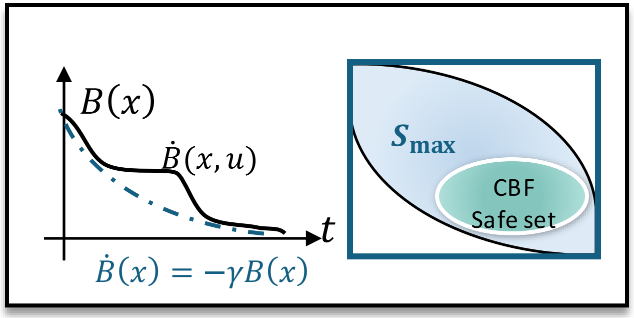


# Background: Control Barrier-Value Function (CBVF) for Safety of Pairwise Interaction

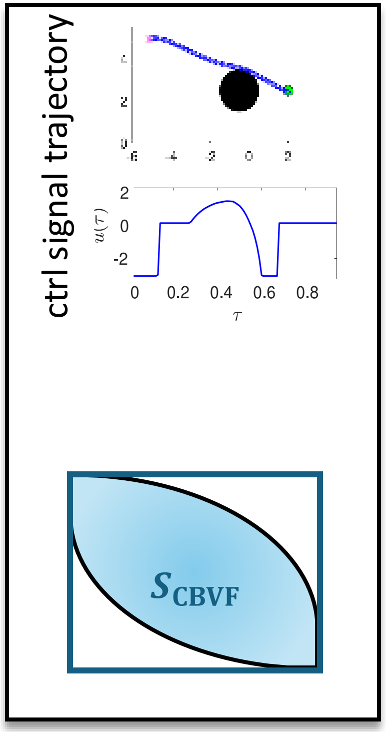
## HJ Reachability

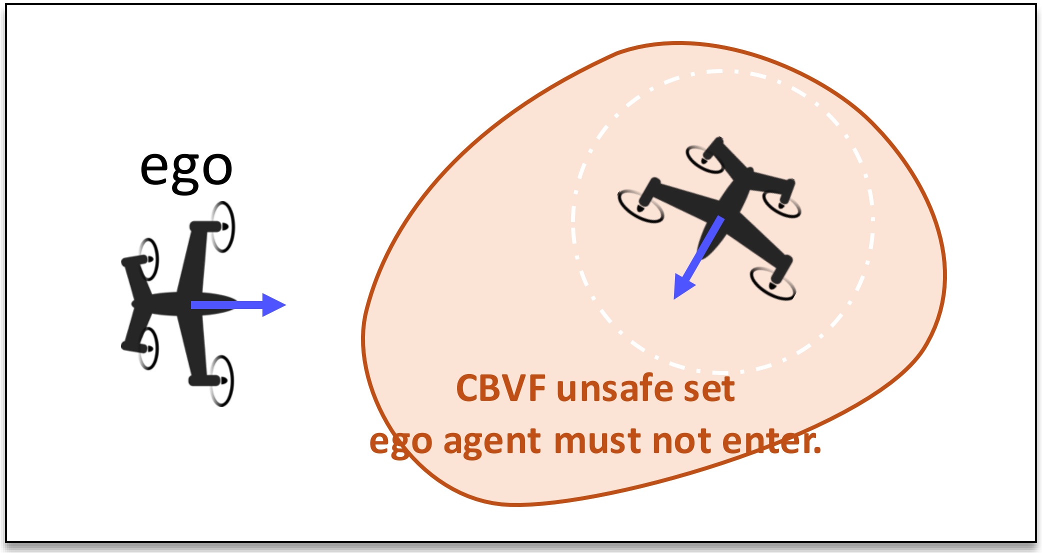



## Control Barrier Functions




## CBVF



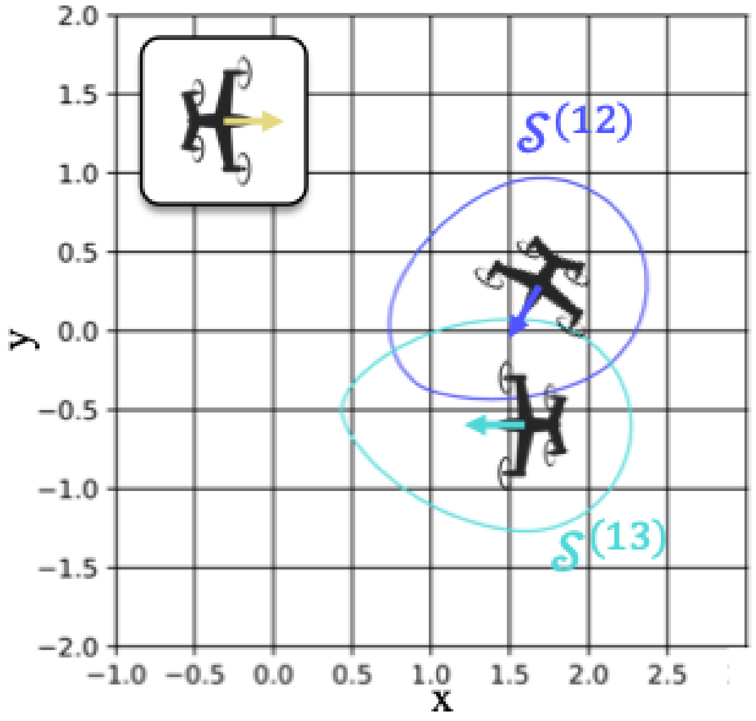



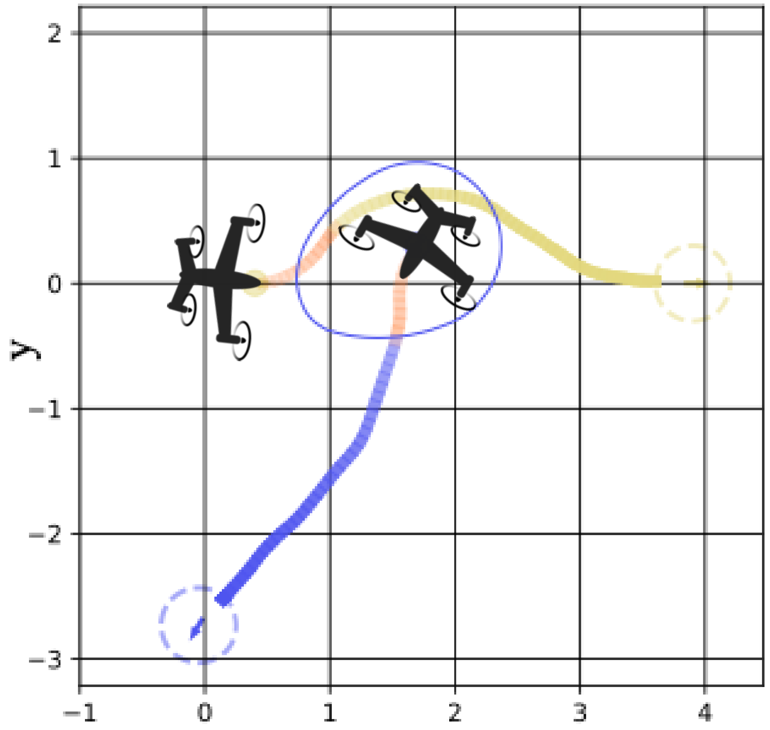

**Choi**, et al., Robust control barrier–value functions for safety-critical control, IEEE CDC'21
**Choi***, **Lee***, et al., A forward reachability perspective on robust control invariance and discount factors in reachability analysis, Arxiv'23 (In prep. for journal)

# How do existing approaches work?

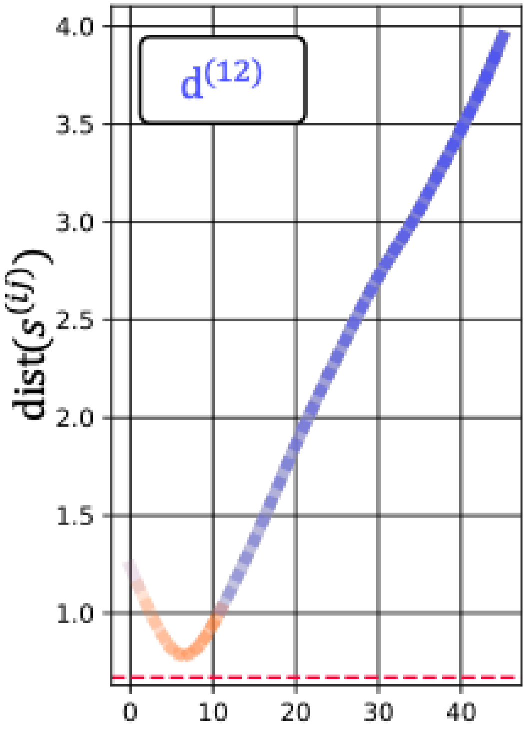
**Traditional Control**: Safety-guaranteed, but struggles with complex multi-agent interactions.

E.g., Two agents



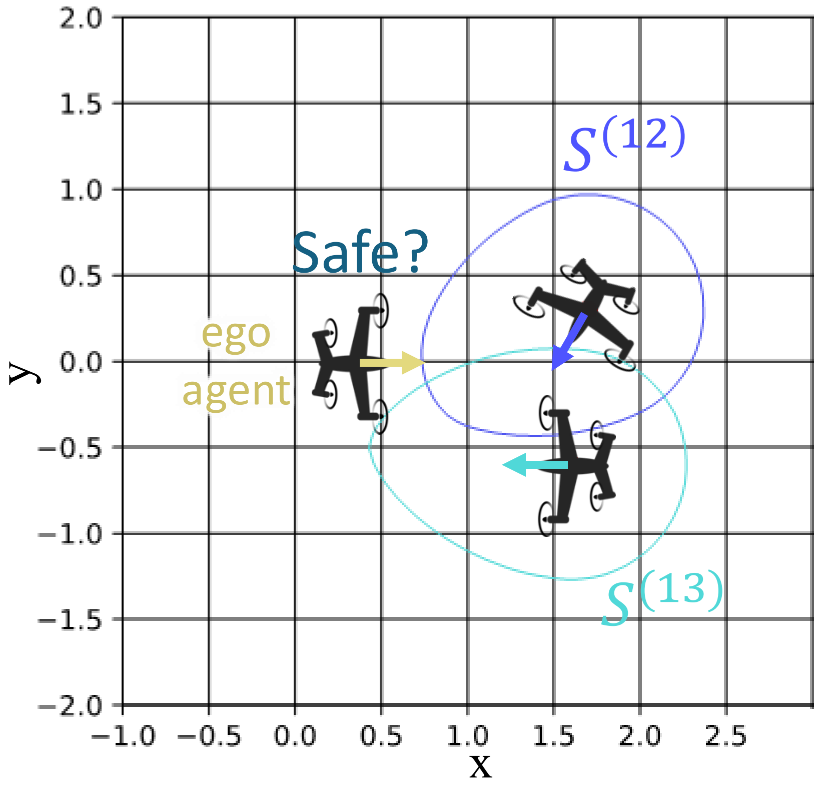




Inter-aircraft distance



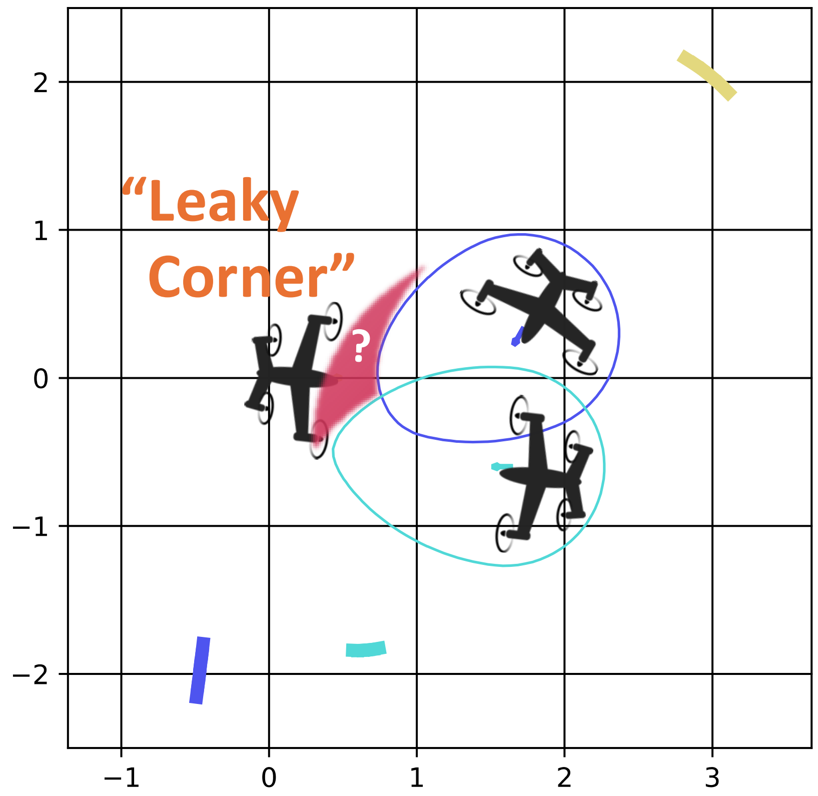

Time

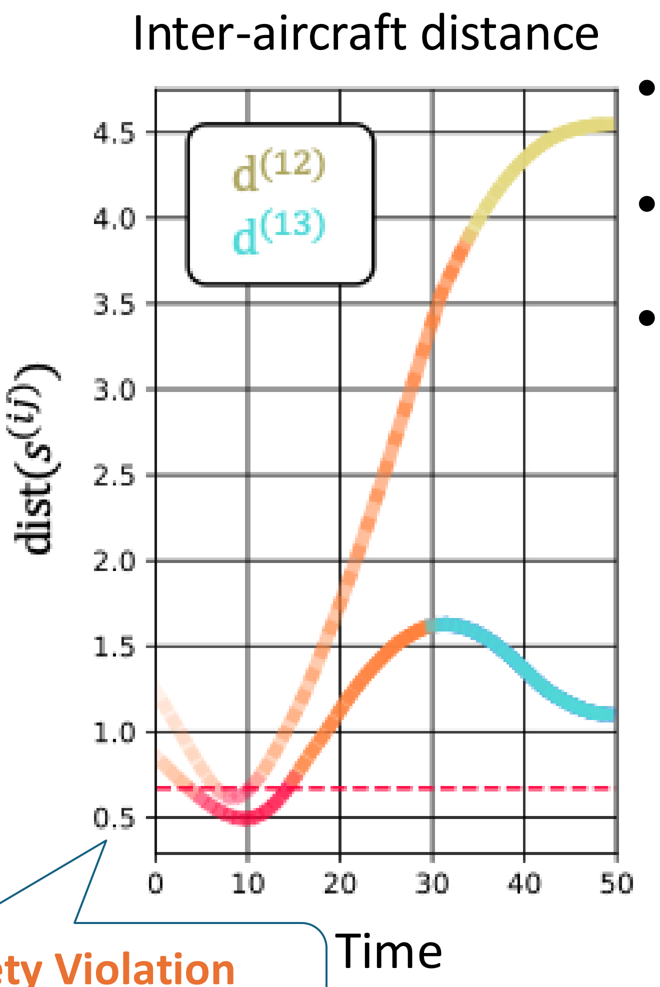
# Now with three agents




: **Max safe set boundaries** from CBVF computation

# Resolving a constraint with two agents may conflict with others




Safety Violation (Near-Collision)

- Creates a gridlock
- Inability to satisfy all constraints
- Collision possible

**Key Challenge**: Intersection of individual safe sets is not the true combined safe sets

Aka "Leaky corner"

$$S(C_1) \cap S(C_2) \neq S(C_1 \cap C_2)$$



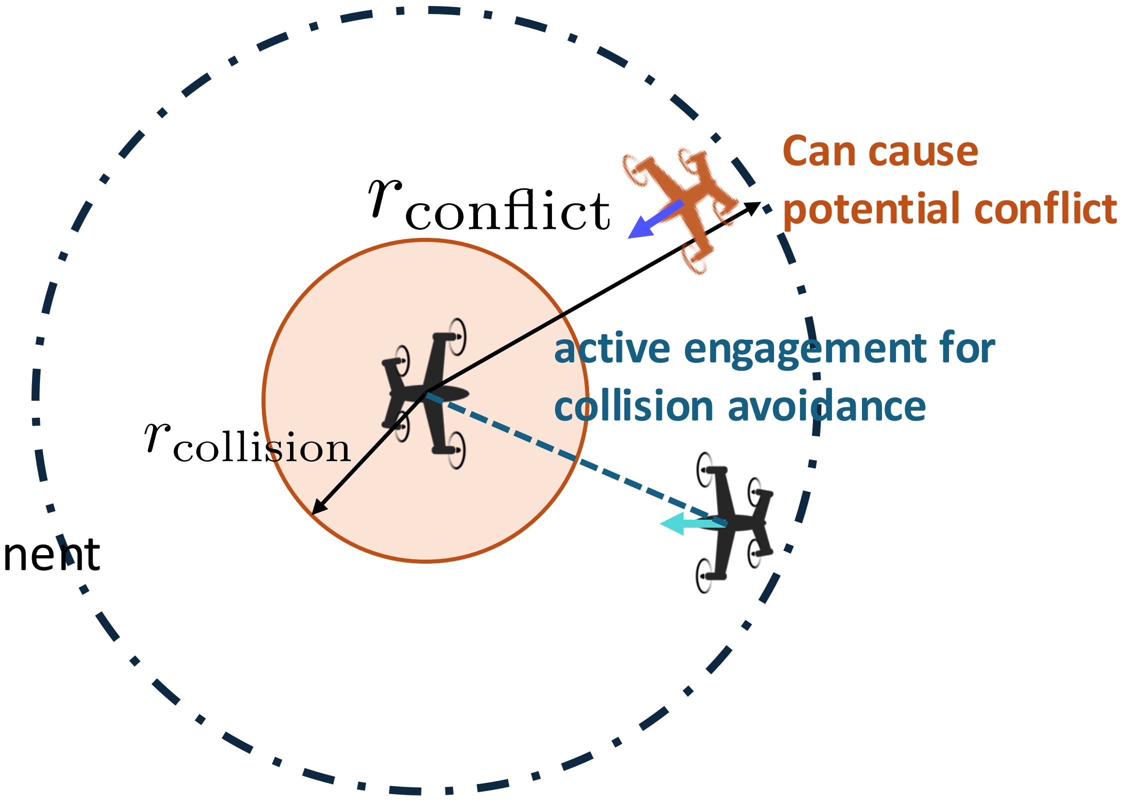
Lee et al., Removing Leaking Corners to Reduce Dimensionality in Hamilton-Jacobi Reachability, ICRA'19

# Distance Ranges Based On Potential Conflict Range

**Definition (informal):**
If more than two vehicles are within the **potential conflict zone**, the leaky corner issue can arise and the safety of all agents is not guaranteed anymore.

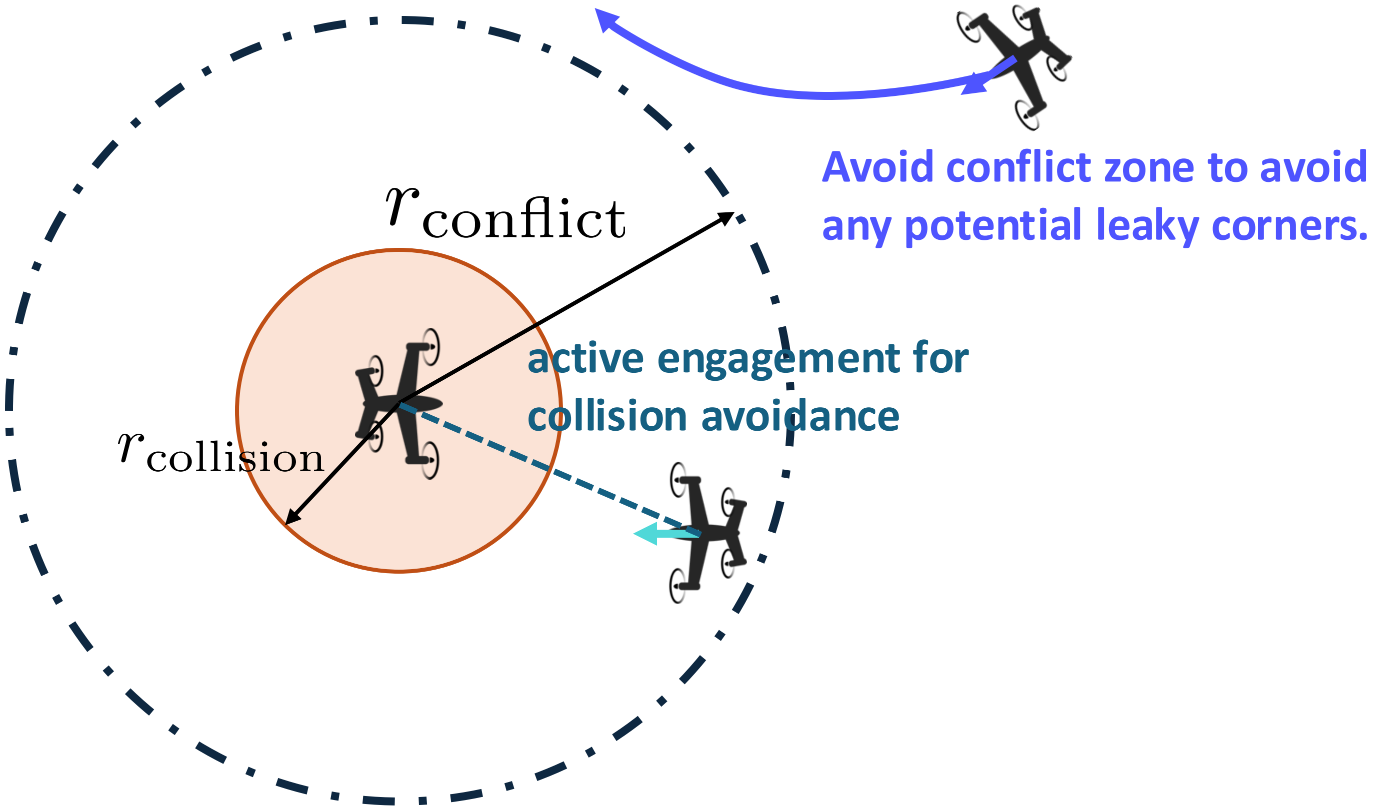
2. $r_{\text{conflict}}$: Potential conflict range

1. $r_{\text{collision}}$: Where collision is imminent



# Distance Ranges Based On Potential Conflict Range

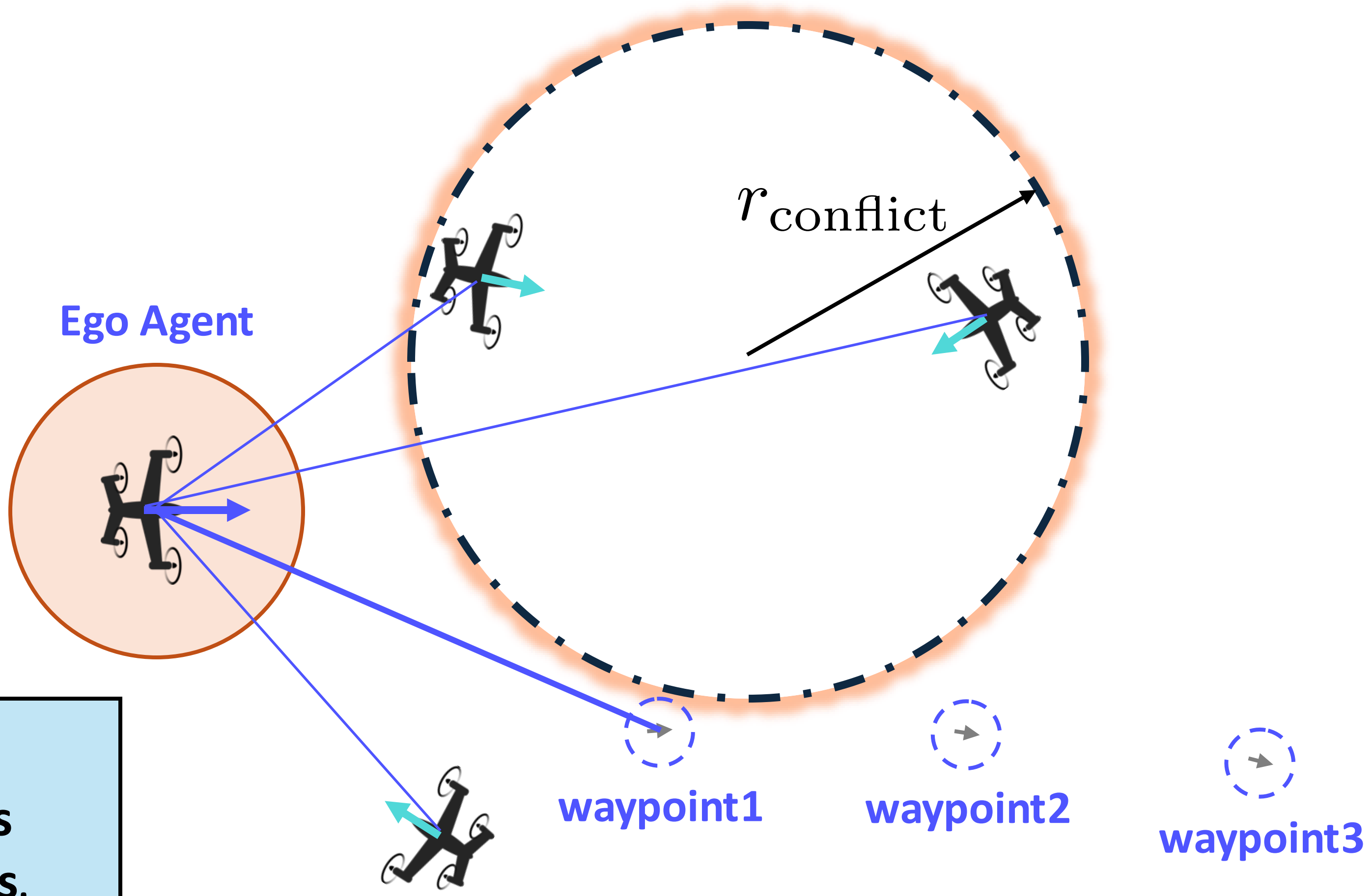
**Definition (informal):**
If more than two vehicles are within the **potential conflict zone**, the leaky corner issue can arise and the safety of all agents is not guaranteed anymore.



# Safety-Informed Multi-Agent RL (MARL)

## Layered Safety for MARL:

**Safe-informed**
Graphical Neural Network
MARL policy




1. Based on the agent's local observations, MARL Policy learns to **navigate to waypoints while avoiding the potential conflict regions**.

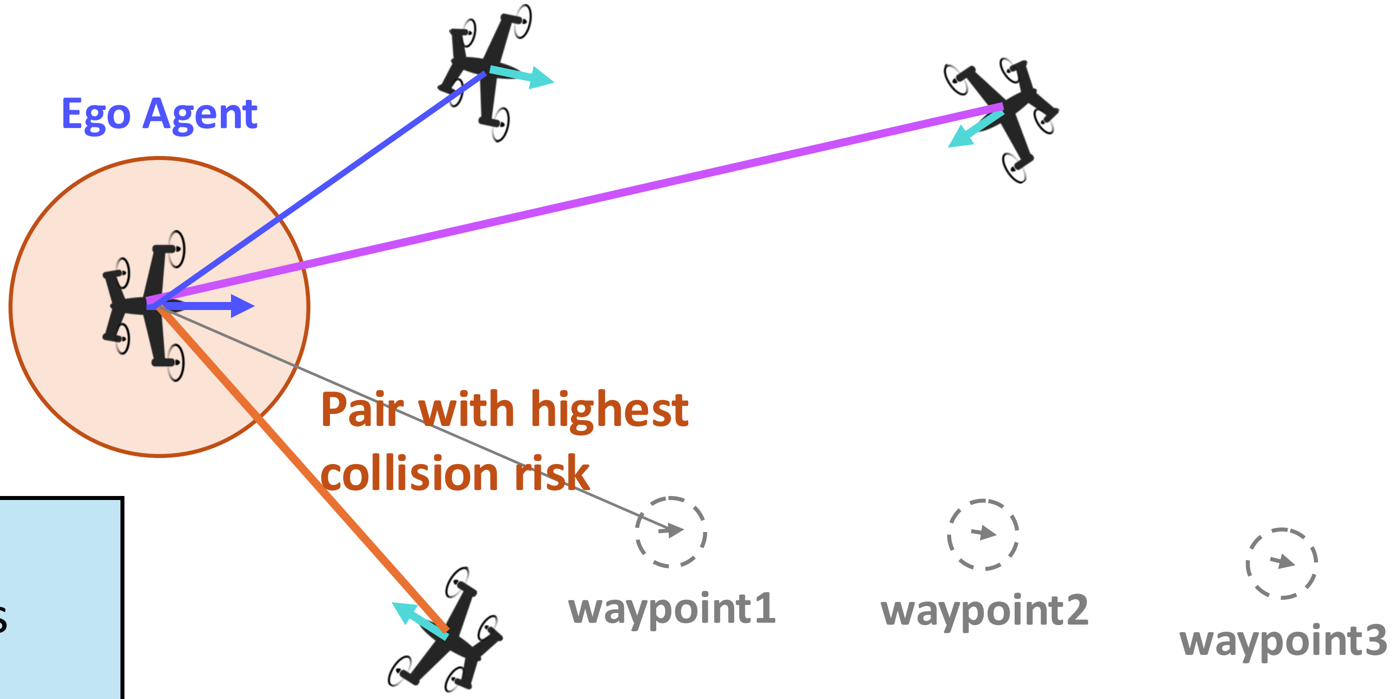
# Safety-Informed Multi-Agent RL (MARL)

## Layered Safety for MARL:

Safe-informed
Graphical Neural Network
MARL policy

**Prioritization**

2. Evaluate CBVF values against each agents and pick the agent with **worst CBVF value**.



# Safety-Informed Multi-Agent RL (MARL)

**Layered Safety for MARL:**

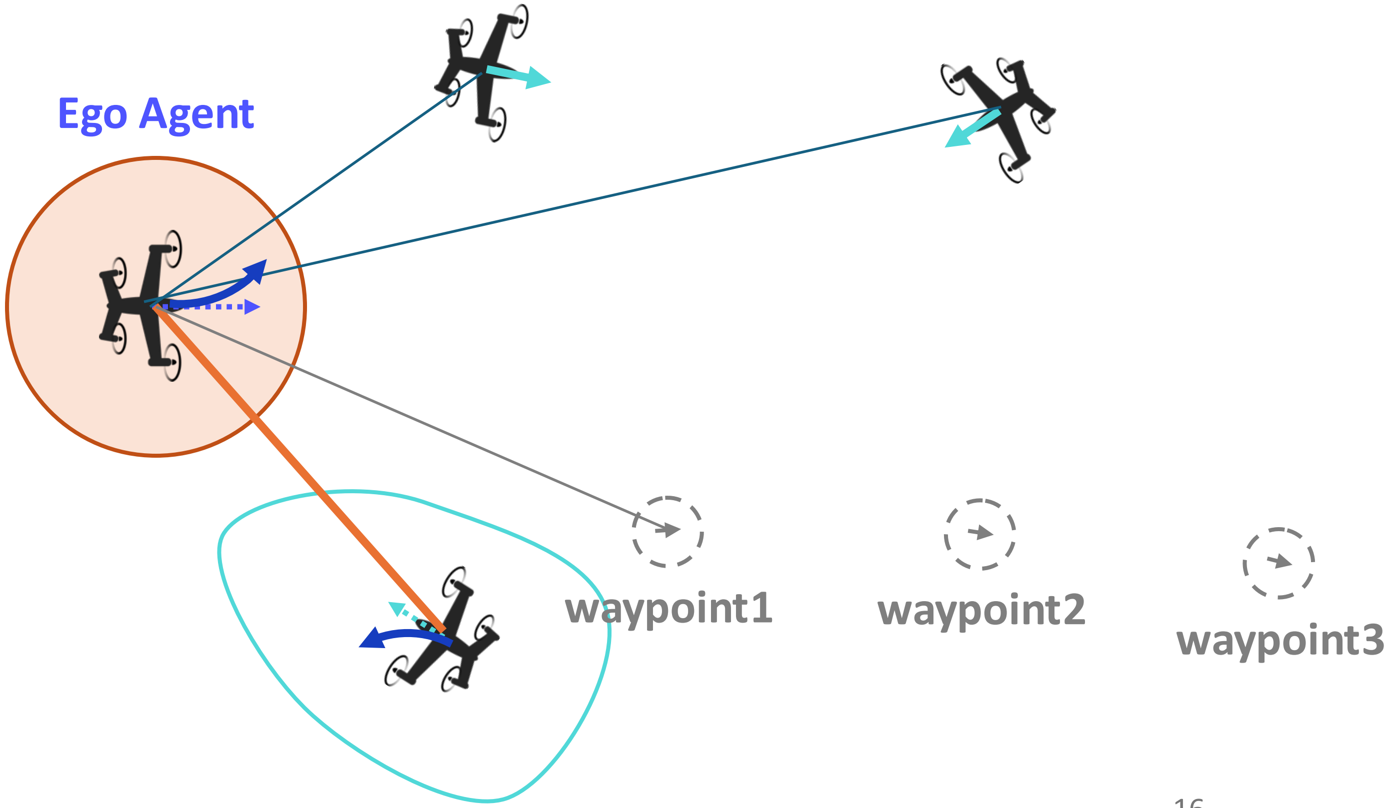
Safe-informed
Graphical Neural Network
MARL policy

↓

Prioritization

↓

**CBVF Safety Filtering**



# Investigating Varying Levels of Safety

Low safety

1. Policy trained without the safety filter and no safety penalty (**safety blind**)

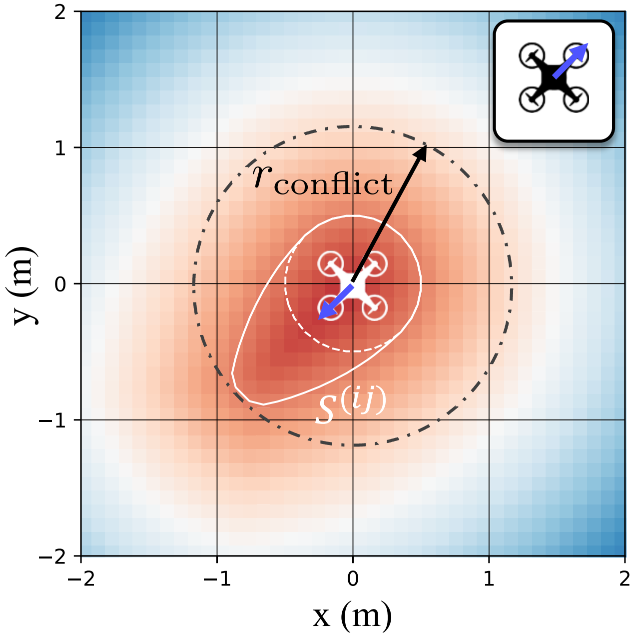
2. Policy trained with the safety filter and no safety penalty (no penalty).

3. Policy trained with the safety filter and with $r_{conflict}$ violation penalty (**proposed**).

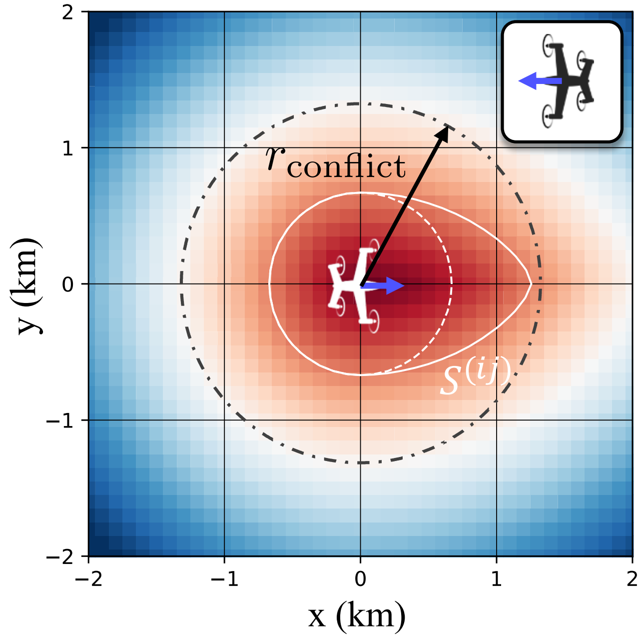
High safety

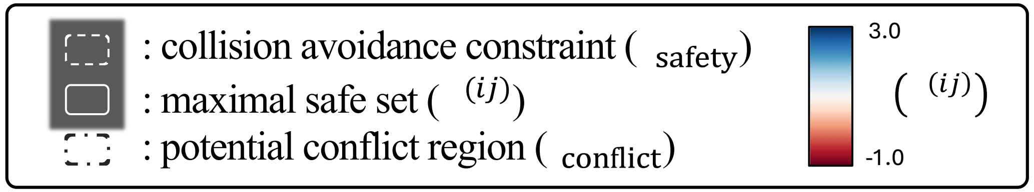
# Conflict zone & safe sets computed with CBVF

**Quadrotor:**
can hover & hold position for safety




**Wingborne Air Taxi:**
Needs active turn for safety ($v_{\min} > 0$)



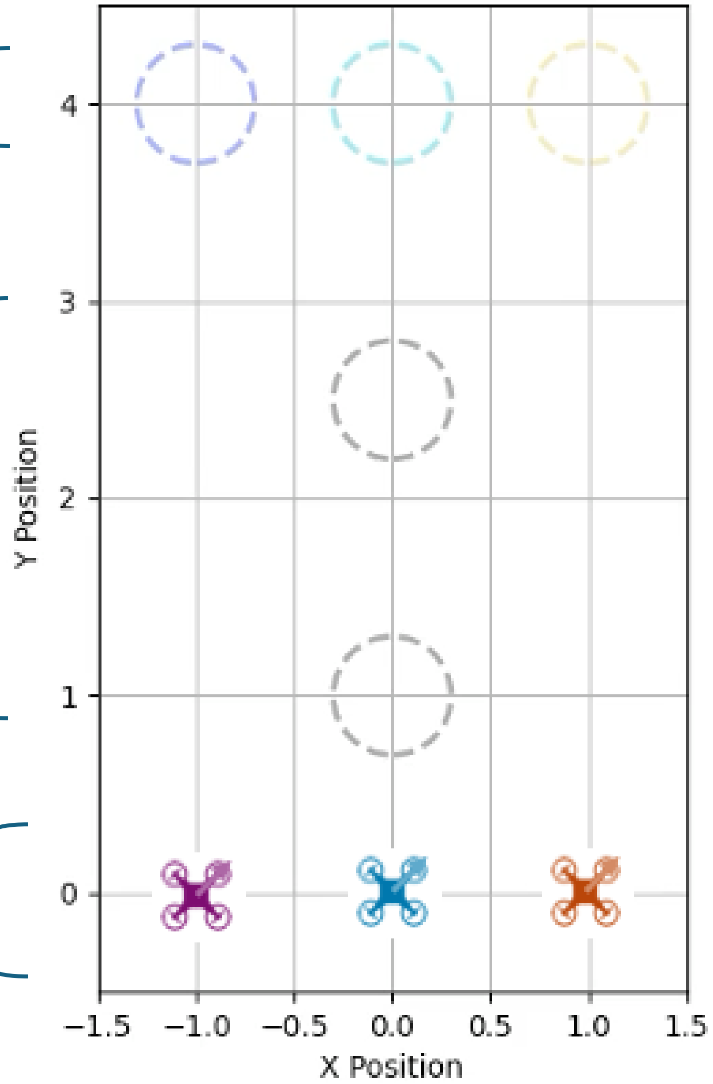



# Safe Navigation Of Quadrotors Through Waypoints

End goal positions

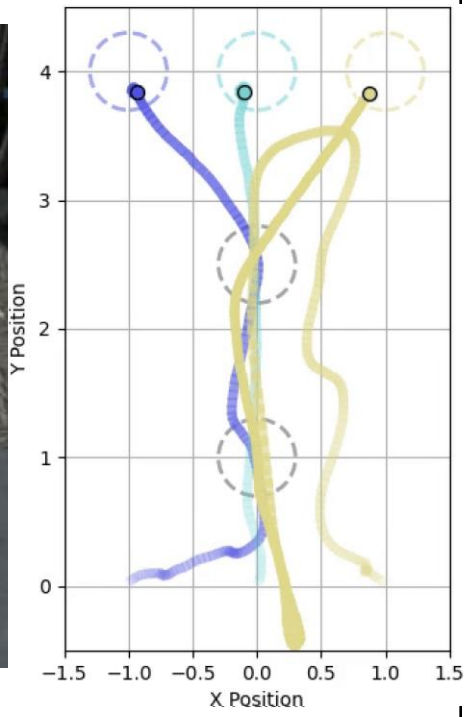
2 waypoints

3 quadrotors 

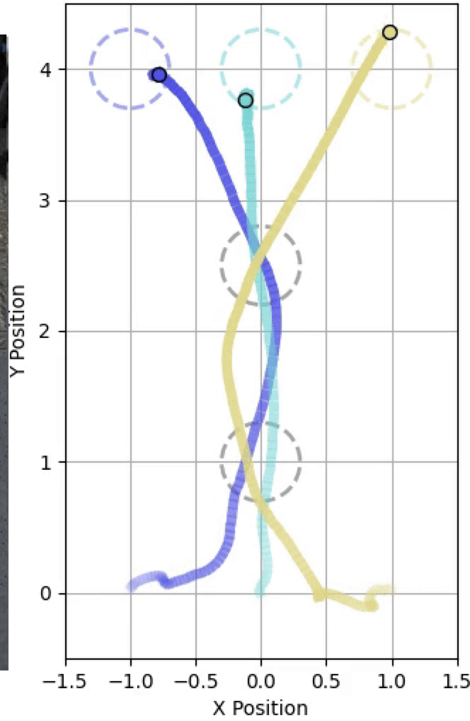


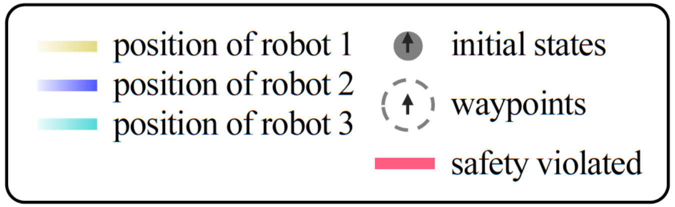
# Safe Navigation Of Quadrotors Through Waypoints

## Safety-blind MARL (safety filtered)




## Safety-informed MARL




2x Speed

# Comparisons to model-based and model-free methods

**Performance**: Goal reach rate (%)

**Safety**: Near collision (%)

## Quadrotor dynamics: N=4 agents

| Methods | Goal reach(%) | Near collision(%) |
|---|---|---|
| DG-PPO | $96 \pm 11.8$ | $0.04 \pm 0.16$ |
| Exponential CBF | $100 \pm 0$ | $0.0 \pm 0.0$ |
| **Our Method** | $100 \pm 0$ | $0.0 \pm 0.0$ |

## Quadrotor dynamics: N=8 agents

| Methods | Goal reach(%) | Near collision(%) |
|---|---|---|
| DG-PPO | $100 \pm 0$ | $9.1 \pm 2.7$ |
| Exponential CBF | $93 \pm 8.9$ | $8.8 \pm 10.7$ |
| **Our Method** | $\mathbf{100 \pm 0}$ | $\mathbf{0.0 \pm 0.0}$ |

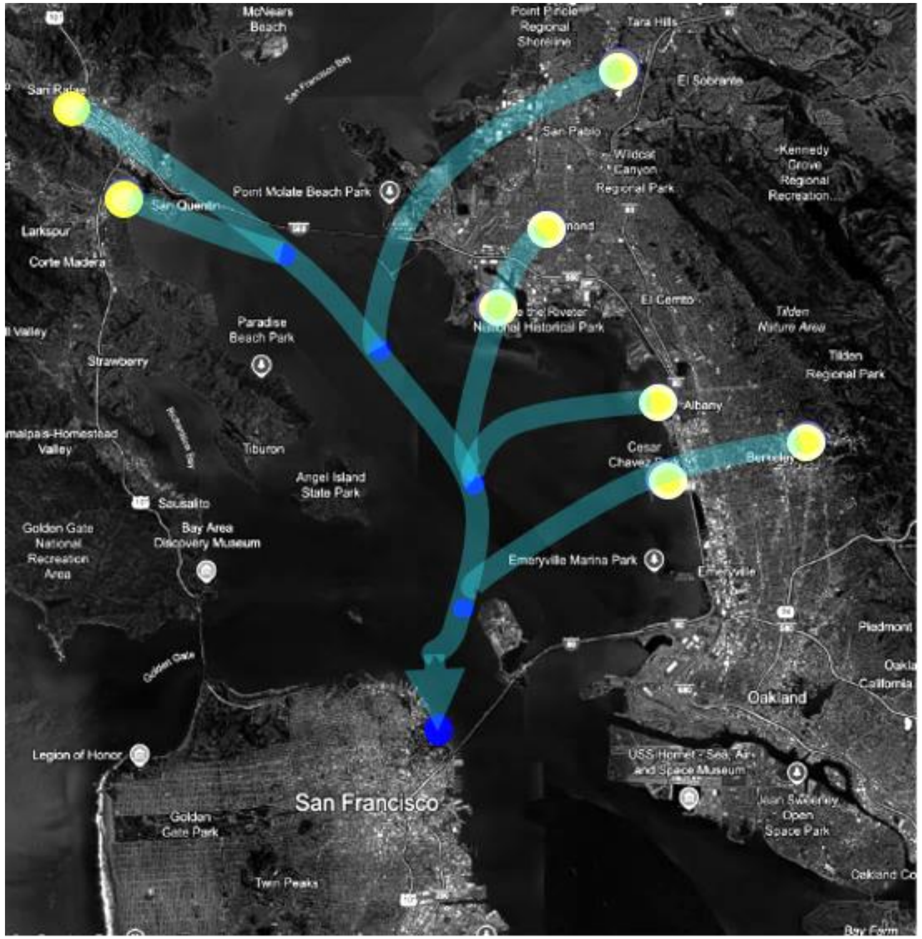
**DG-PPO**: Songyuan Zhang, Oswin So, Mitchell Black, and Chuchu Fan. Discrete GCBF Proximal Policy Optimization for Multi-agent Safe Optimal Control. ICLR 2025.
**Exponential CBF**: Quan Nguyen and Koushil Sreenath. Exponential control barrier functions for enforcing high relative-degree safety-critical constraints. ACC, 2016.

# Autonomous Urban Air Mobility Traffic Management: Air Taxi Operation Scenario

**Urban Air Mobility (UAM) demand analysis**

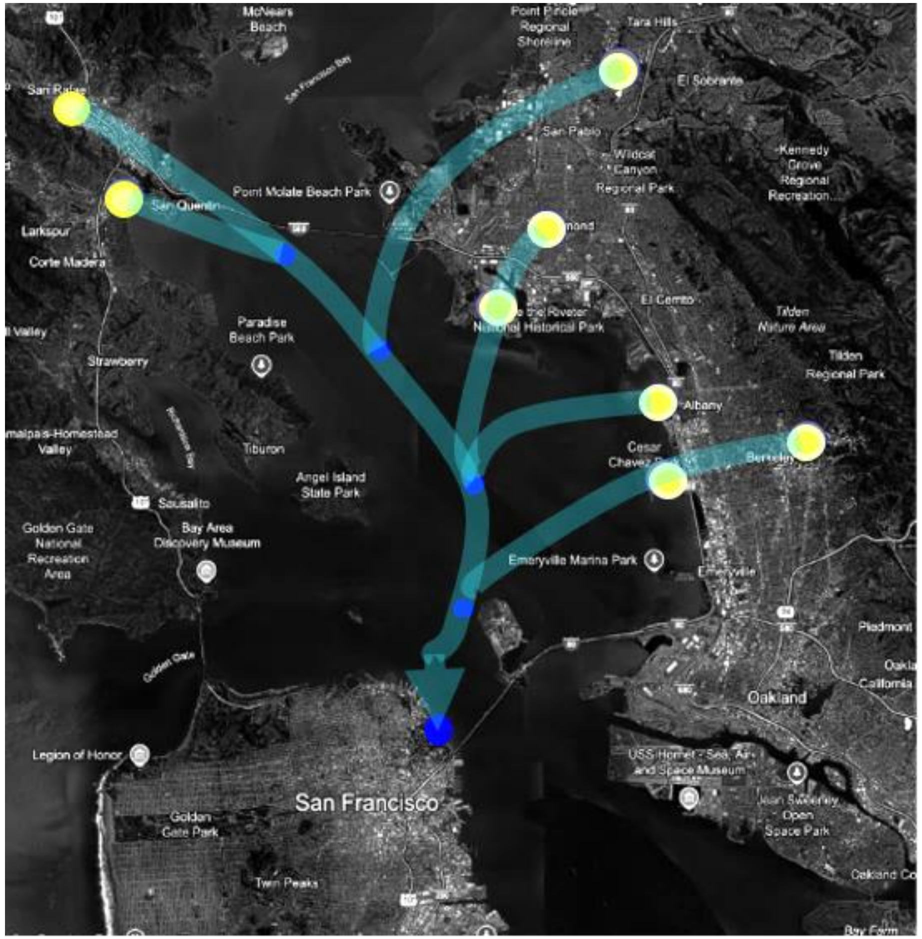
Consider a peak-density scenario:

- Each vertiport serves 125 aircraft/hour
- Two operations (takeoffs and landings)/min,
- Corridors with 1500 ft separation




(8 vehicles merging into the air corridor that lands at San Francisco)

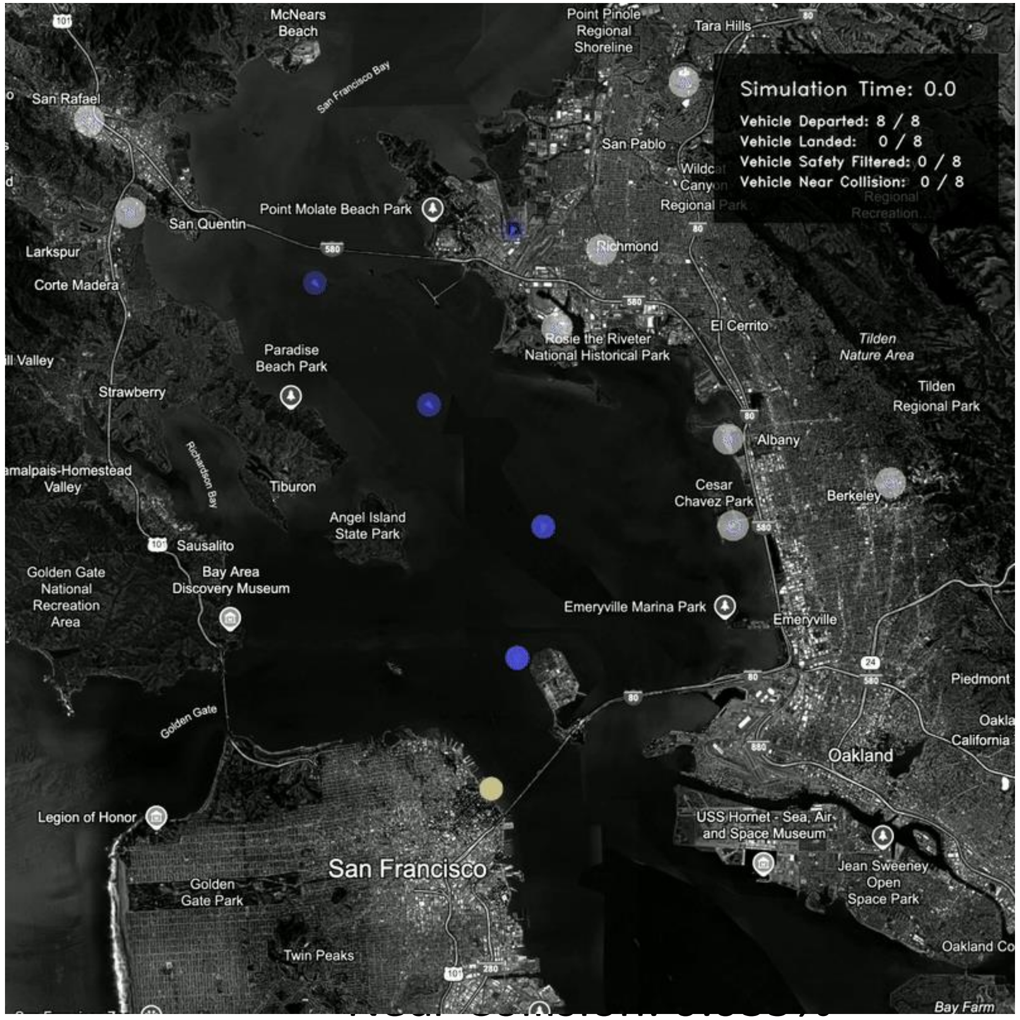
# Autonomous Urban Air Mobility Traffic Management: Air Taxi Operation Scenario



(8 vehicles merging into the air corridor that lands at San Francisco)

# Air Taxi Operation Scenario: Merging into corridor

## Safety-blind MARL (safety filtered)

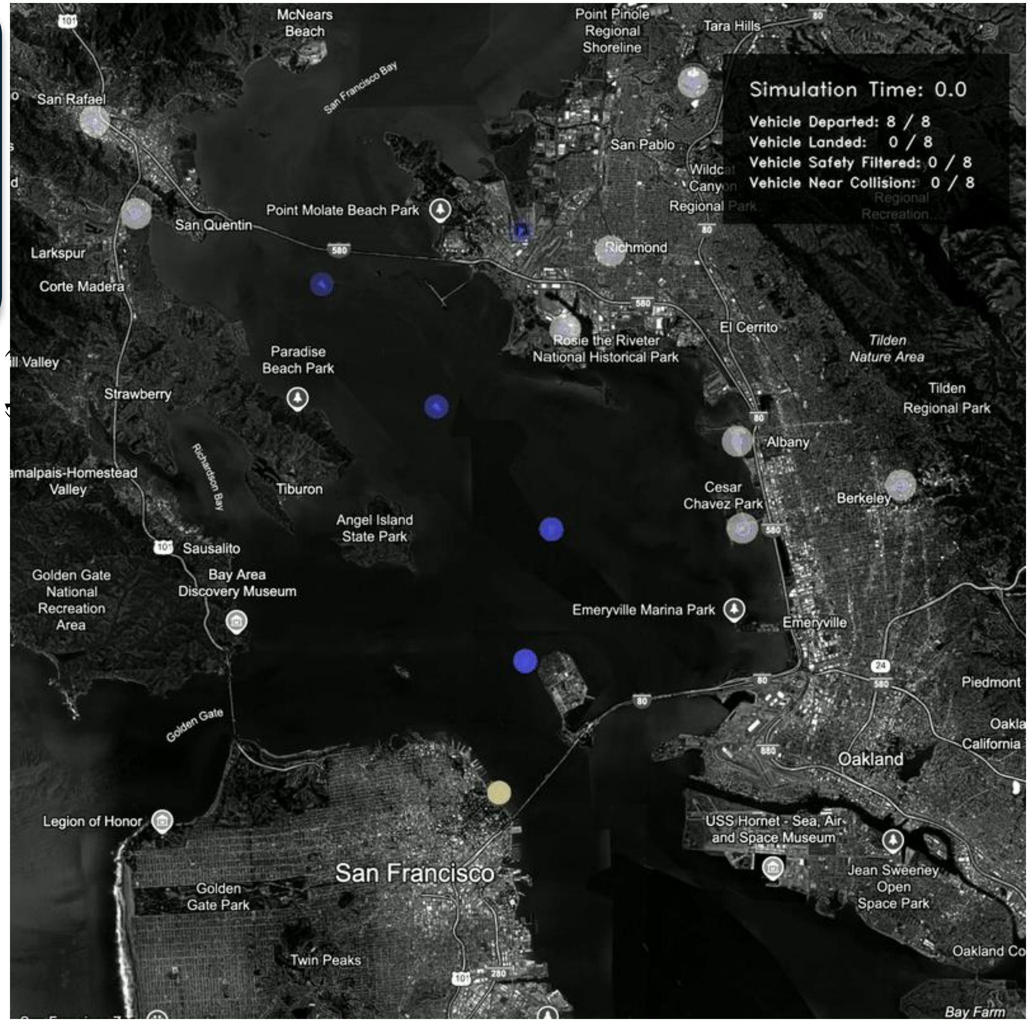



position of vehicles

Active CBVF-based safety filtering

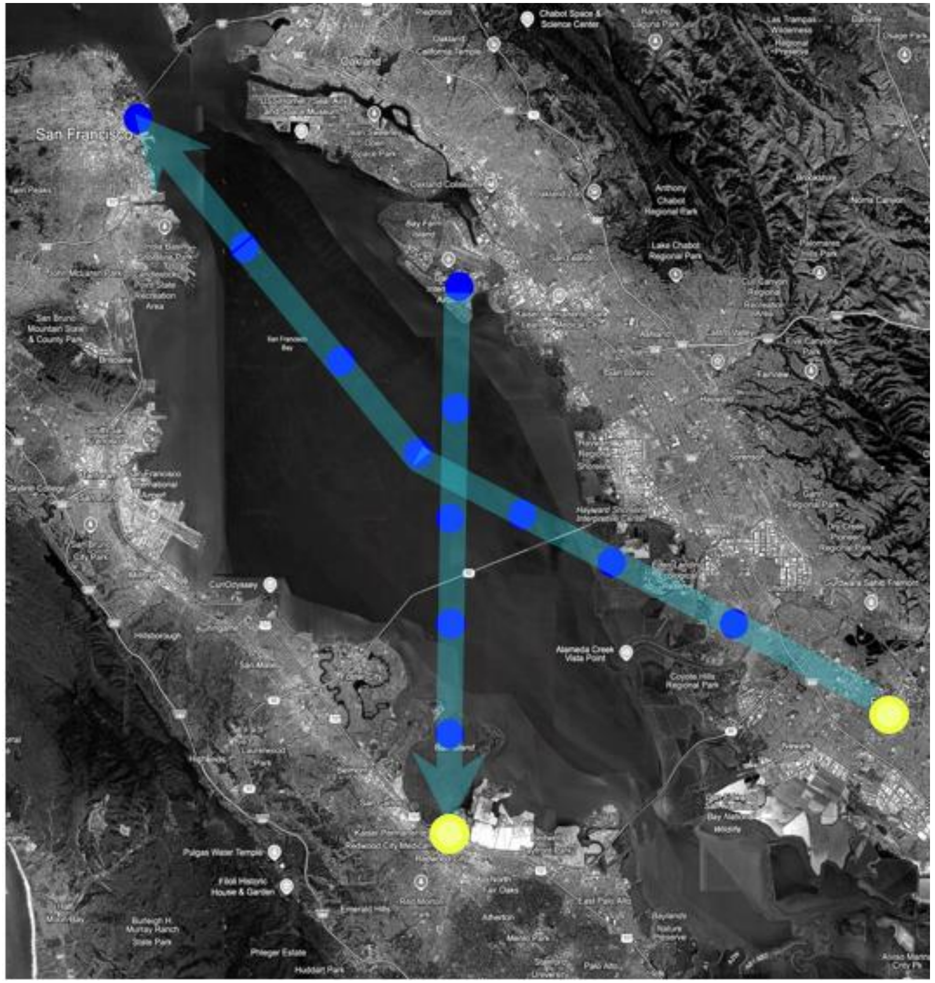
10x Speed

## Safety-informed MARL



# Air Taxi Operation Scenario: Intersecting corridors

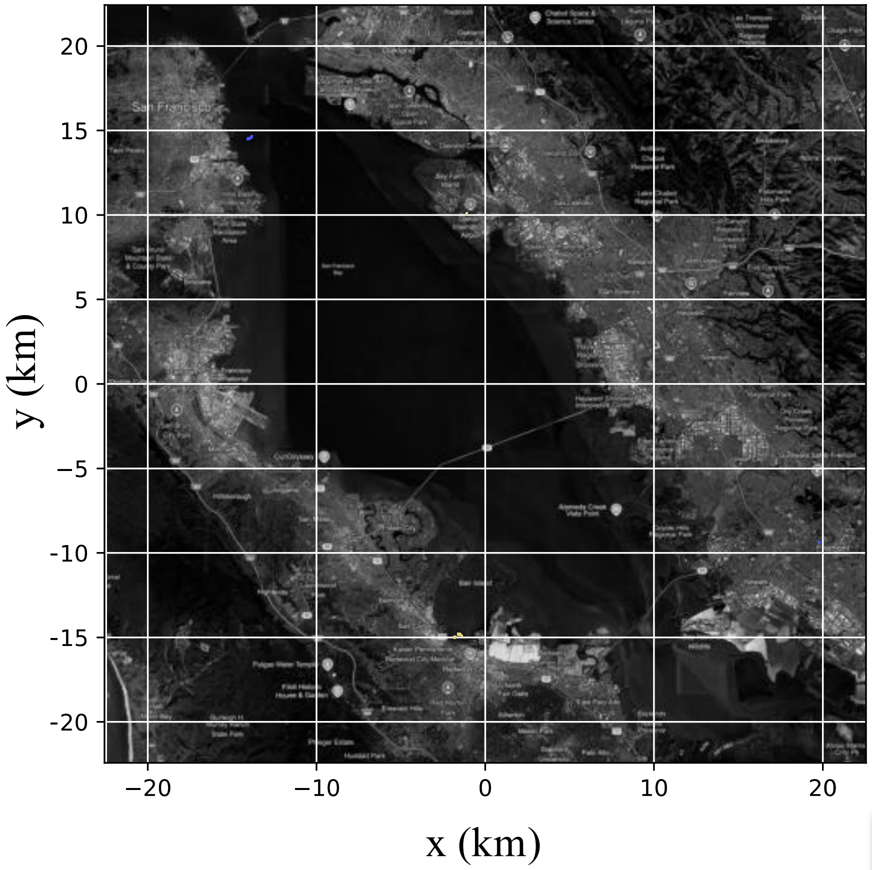
**Intersection Scenario**




8 vehicles in each air corridor (total 16), crossing each other

# Urban Air Mobility Simulation Result: Intersection Scenario

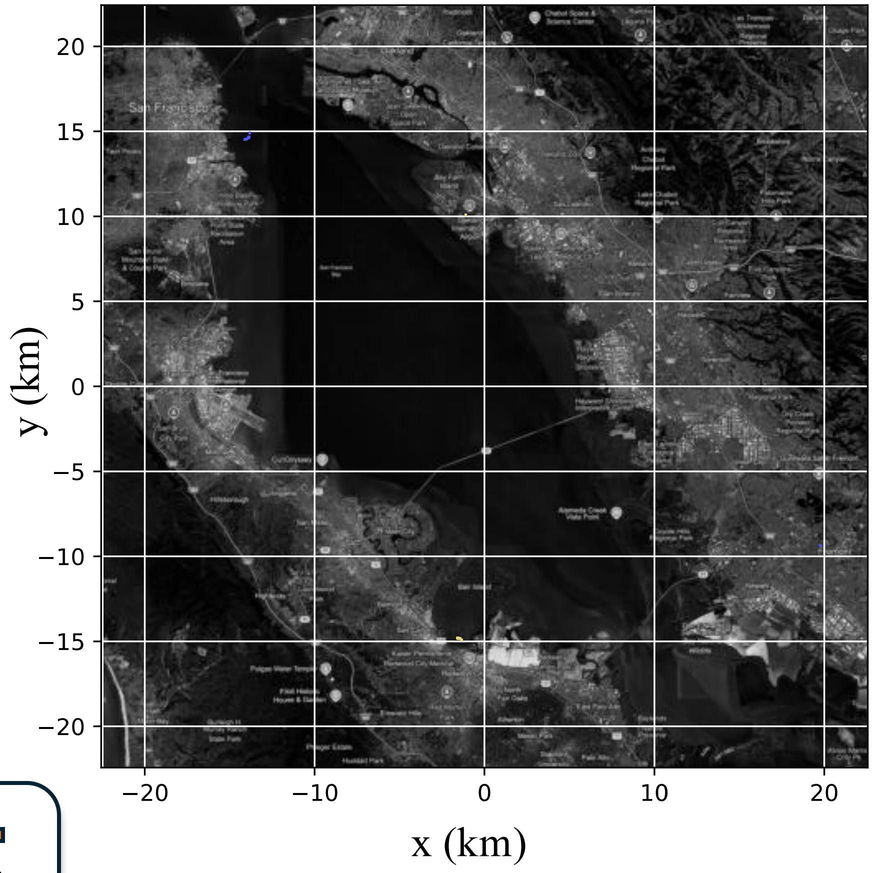
Safety-blind MARL (safety filtered)




Avg. Travel Time: 781 sec
Near Collision: 0.129%

Active CBVF-based safety filtering

Safety-informed MARL




Avg. Travel Time: 660.8 sec
Near Collision: 0.056%

# Resolving Conflicting Constraints in MARL with Layered Safety

- **Layered architecture:** Combines CBVF safety filtering with MARL policy for safe, efficient navigation.

- **Proven performance:** Faster travel, more waypoints covered, fewer conflicts.

- **Broad evaluation:** Tested on quadrotor and fixed-wing dynamics in complex scenarios.

# Conclusions

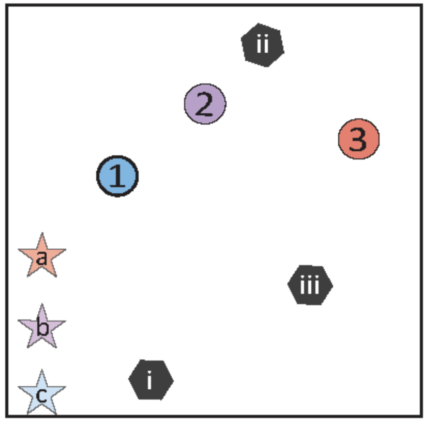
**Implications of this work:**

- ✓ Our method integrates model-based safety tools from control theory (CBVFs) with learning-based methods (MARL),
- ✓ Together forming a framework that addresses two major challenges in multi-agent problems—safety and efficient coordination.
- ✓ Our work demonstrate the viability of hybrid approaches that combine learning and control and illustrate how RL can be responsibly applied in safety-critical settings.

# Proposed Layered Safety Architecture

## Multi-agent reinforcement learning

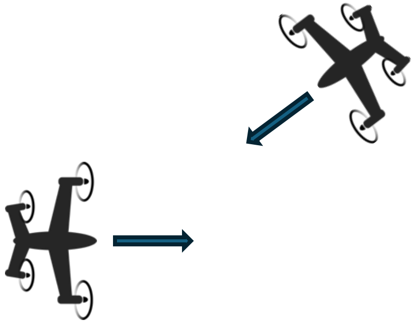
- Learns to proactively navigate potentially conflicting constraints
- Optimizes task performance (e.g., go to goal)

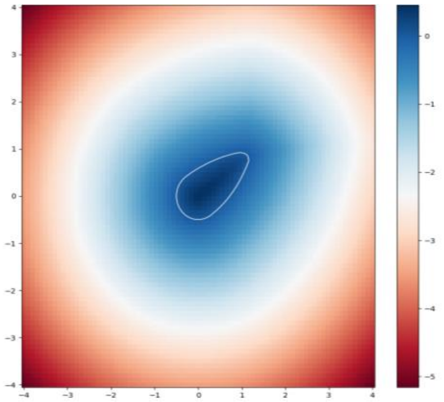



## Control Barrier Value Functions (CBVFs)

a CBF design method based on Hamilton-Jacob reachability

- Filters unsafe action generated by MARL
- Provides a safety-informed reward signal





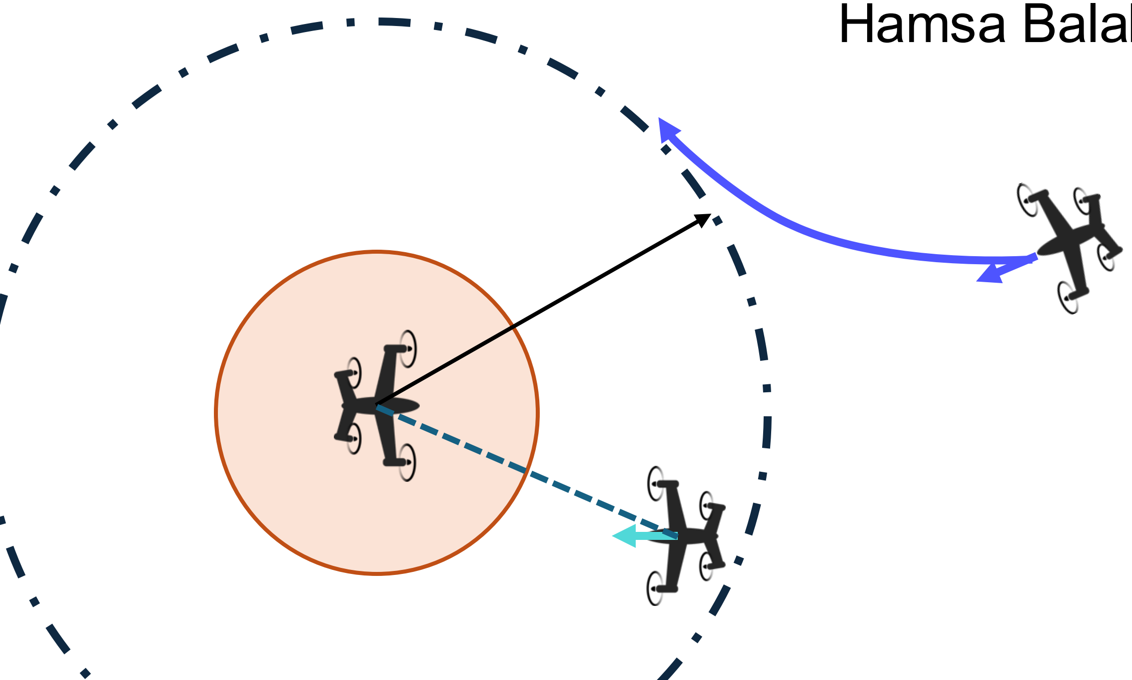
**Layered safety architecture**

enhances both safety and performance

# For more details check out our RSS 2025 Paper!

## Resolving Conflicting Constraints in Multi-Agent Reinforcement Learning with Layered Safety

Jason J. Choi*, Jasmine Jerry Aloor*, Jingqi Li*, Maria G. Mendoza,
Hamsa Balakrishnan, Claire J. Tomlin



Webpage, Paper, Videos

Hybrid Systems Laboratory

DINaMo