LEARNING LARGE SCALE COMPETITIVE TEAM BEHAVIORS WITH MEAN-FIELD INTERACTIONS

Bhavini Jeloka, Yue Guan, Panagiotis Tsiotras

School of Aerospace Engineering

Georgia Institute of Technology

Georgia Tech DCSL

Manuscript

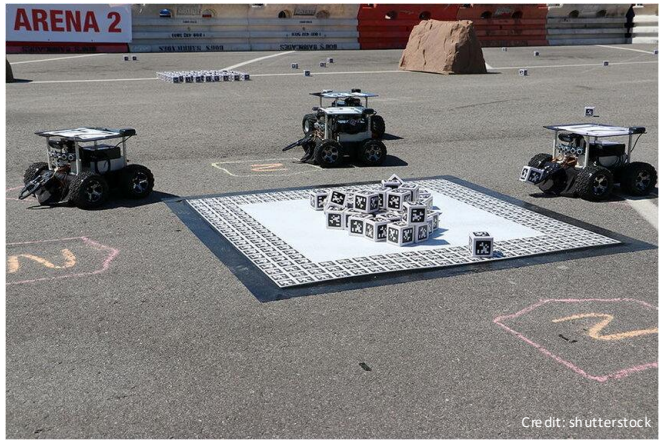
NEURAL INFORMATION PROCESSING SYSTEMS

Motivation: Large-Scale Multi-Agent Interactions

Mixed Collaborative-Competitive Settings

- Team-level: Competition
- Within each team: Cooperation

Swarm Systems



Sports



Battlefield Offence-Defense



Key Challenges

Scalability
(curse of dimensionality)

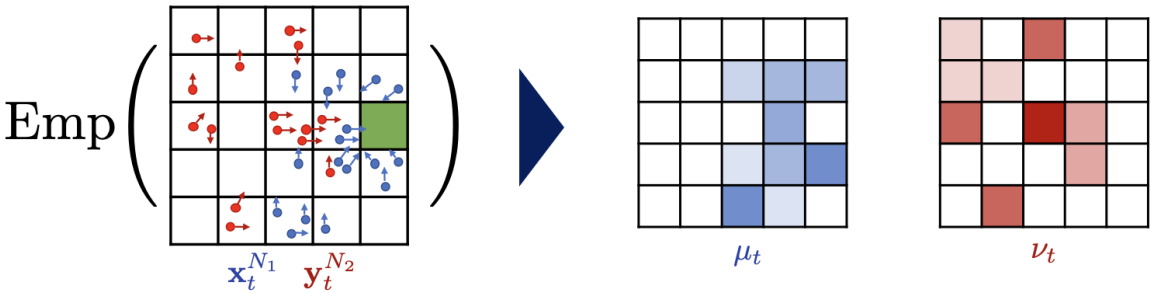
Information Structure
(what does each agent know?)

Opponent Modelling
(what is my competitor's plan?)

Zero-Sum Mean-Field Team Games

- Game between two teams: **Blue** and **Red**
- N_1 and N_2 Homogenous Agents
- Finite State and Action Spaces (MDP-type)

Empirical Distribution ~ fraction of agents in each state

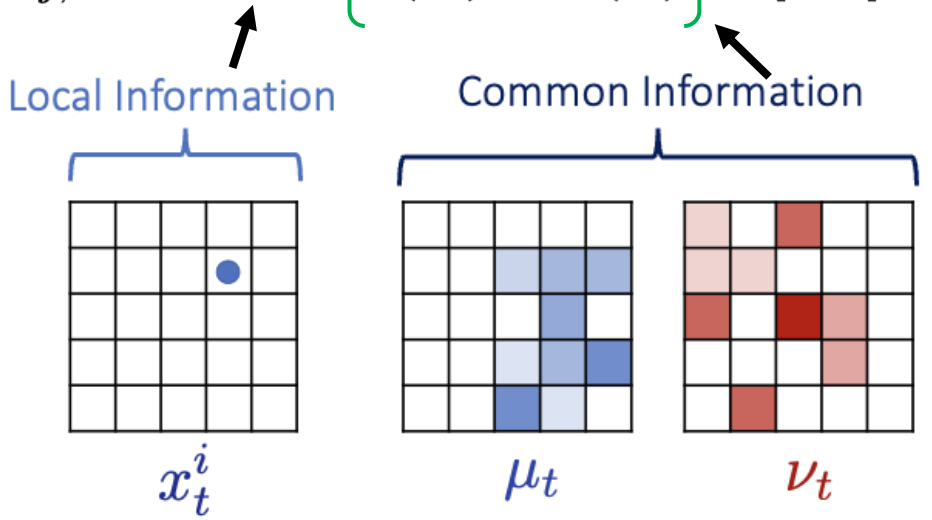


Weakly coupled transition dynamics and rewards (in ED)

ED at infinite population limit: **Mean-Field!**

(Stochastic) Policies under Mean-Field Sharing

$$\phi_{i,t} : \mathcal{U} \times \mathcal{X} \times \{\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})\} \to [0,1]$$
$$\psi_{j,t} : \mathcal{V} \times \mathcal{Y} \times \{\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})\} \to [0,1],$$



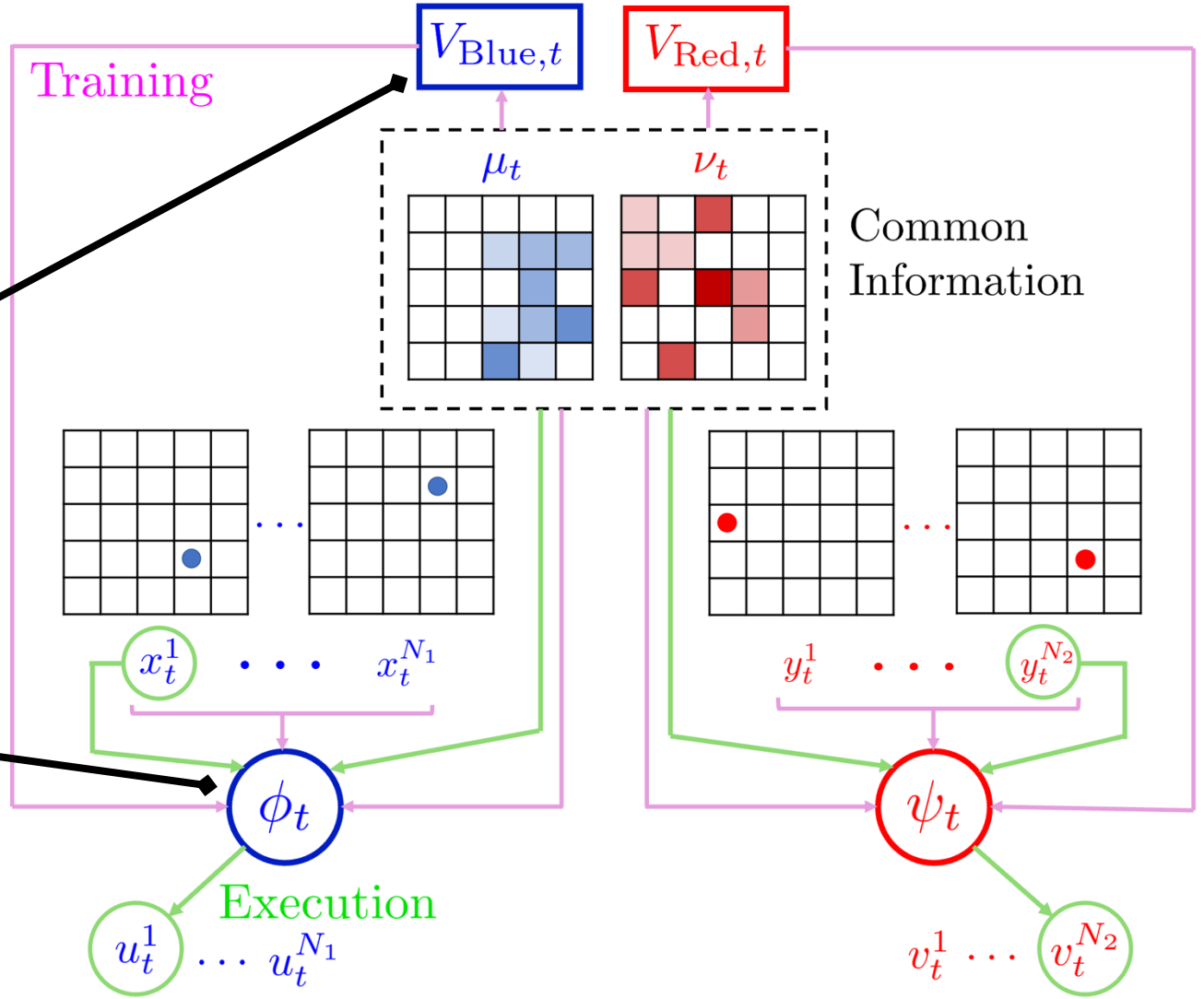
Zero-sum: **Blue** maximizes & **Red** minimizes

Ref: Guan, Yue, Mohammad Afshari, and Panagiotis Tsiotras. "Zero-sum games between mean-field teams: Reachability-based analysis under mean-field sharing." (AAAI 2024)

Mean-Field Multi-Agent Proximal Policy Optimization

✓ **Minimally-Informed Critic**
- Only requires common information (EDs) – smaller network size
- Scale with # of agents: ✗

✓ **Shared-Team Actor**
- Identical team policies (Theorem 3.1)
- Only EDs + private information enters *individually*



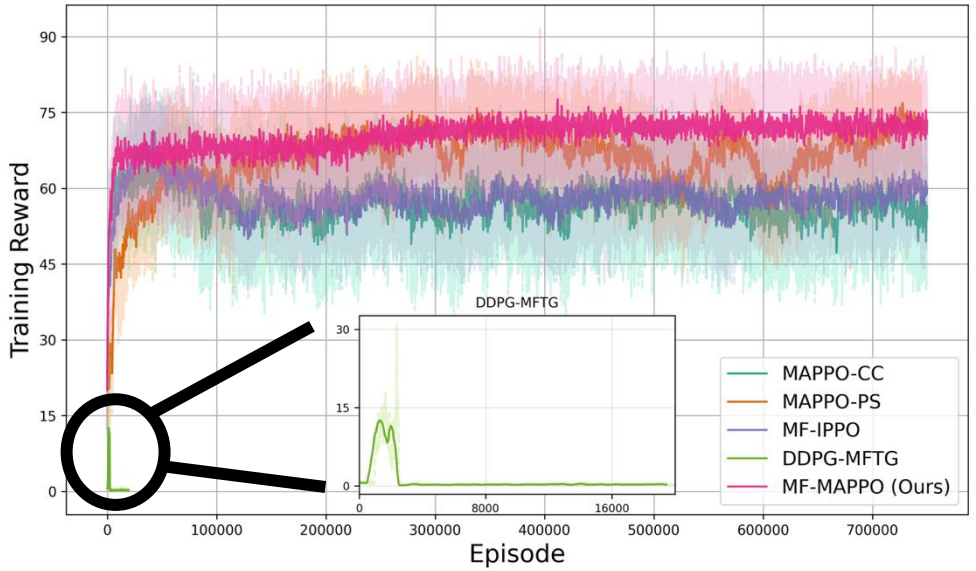
✓ **Finite-population Training**
No MF *oracle*

✓ **Memory Efficient**
One buffer team

✓ **Scalable**
Train on N and deploy on $M > N$ (Theorem 3.3)

Numerical Experiments: Battlefield

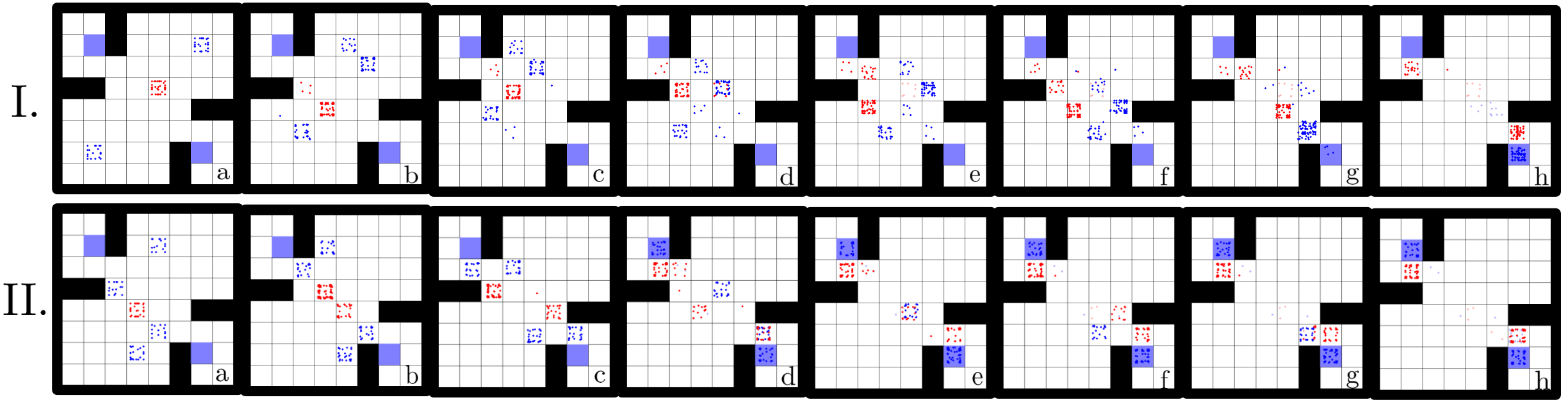
- Battlefield on 2D grid world
  - Blue team: reach the target location(s)
  - Red team: defend target(s)
- Reward $\propto$ fraction of Blue population at target
- Dynamics: Deactivated by the opponent team's agents based on relative numerical advantage



✓ OUTPERFORM BASELINES

Battlefield	MF-IPPO	MAPPO-PS	MAPPO-CC	DDPG-MFTG	MF-MAPPO (Ours)
MF-IPPO	61.54	73.85	70.60	0.00	**86.26**
MAPPO-PS	55.82	72.78	71.64	0.00	**86.22**
MAPPO-CC	51.60	68.81	65.58	0.00	**82.58**
DDPG-MFTG	65.26	75.14	77.21	0.00	**85.95**
MF-MAPPO (Ours)	**36.10**	**54.43**	**62.00**	0.00	**76.26**

Numerical Experiments: Battlefield



Teams are equally split amongst occupied cells

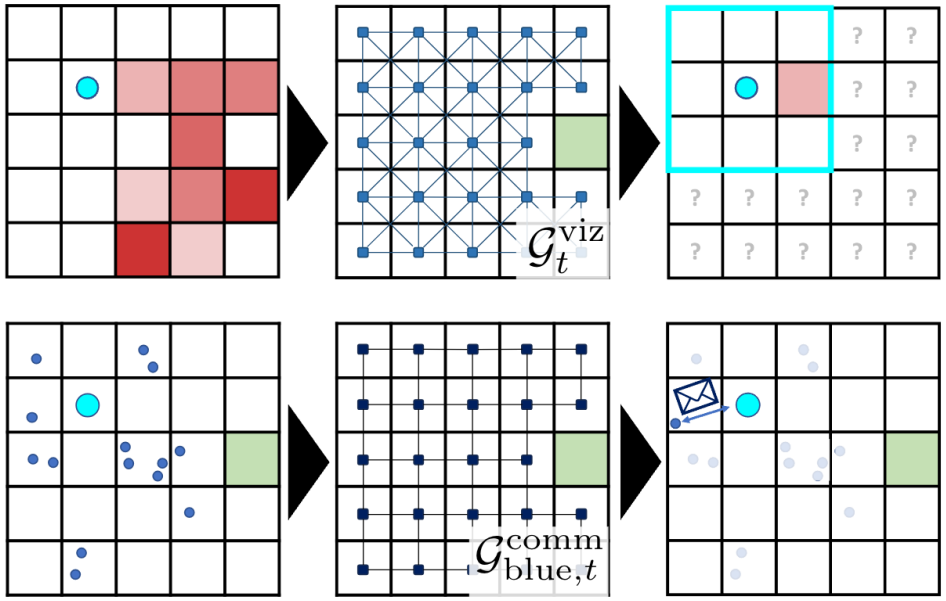
✓ **HETEROGENEOUS BEHAVIOR**

Agents show intelligent, heterogeneous behavior in **attacking/defending** targets despite identical team policies

Partially-Observable Mean-Field Team Games

Dynamic-Projected Consensus

Estimating MF using local observations and communication graph (Theorem B.1)



Gradient-Regularized MF-MAPPO

$$L_{\text{actor}}^{\text{GR}}(\theta_{\text{Blue}}) = L_{\text{actor}}(\theta_{\text{Blue}}) + \frac{\lambda^{\text{GR}}}{|\mathcal{D}|} \sum_{\mathcal{D}} \left\| \nabla_{\mu,\nu} \log \phi(u|x,\mu,\nu) \right\|_2^2$$

Lipschitz Policies: Desensitized to small errors in mean-field estimation.

