LEARNING LARGE-POPULATION COMPETITIVE BEHAVIORS

A MEAN-FIELD TEAM APPROACH

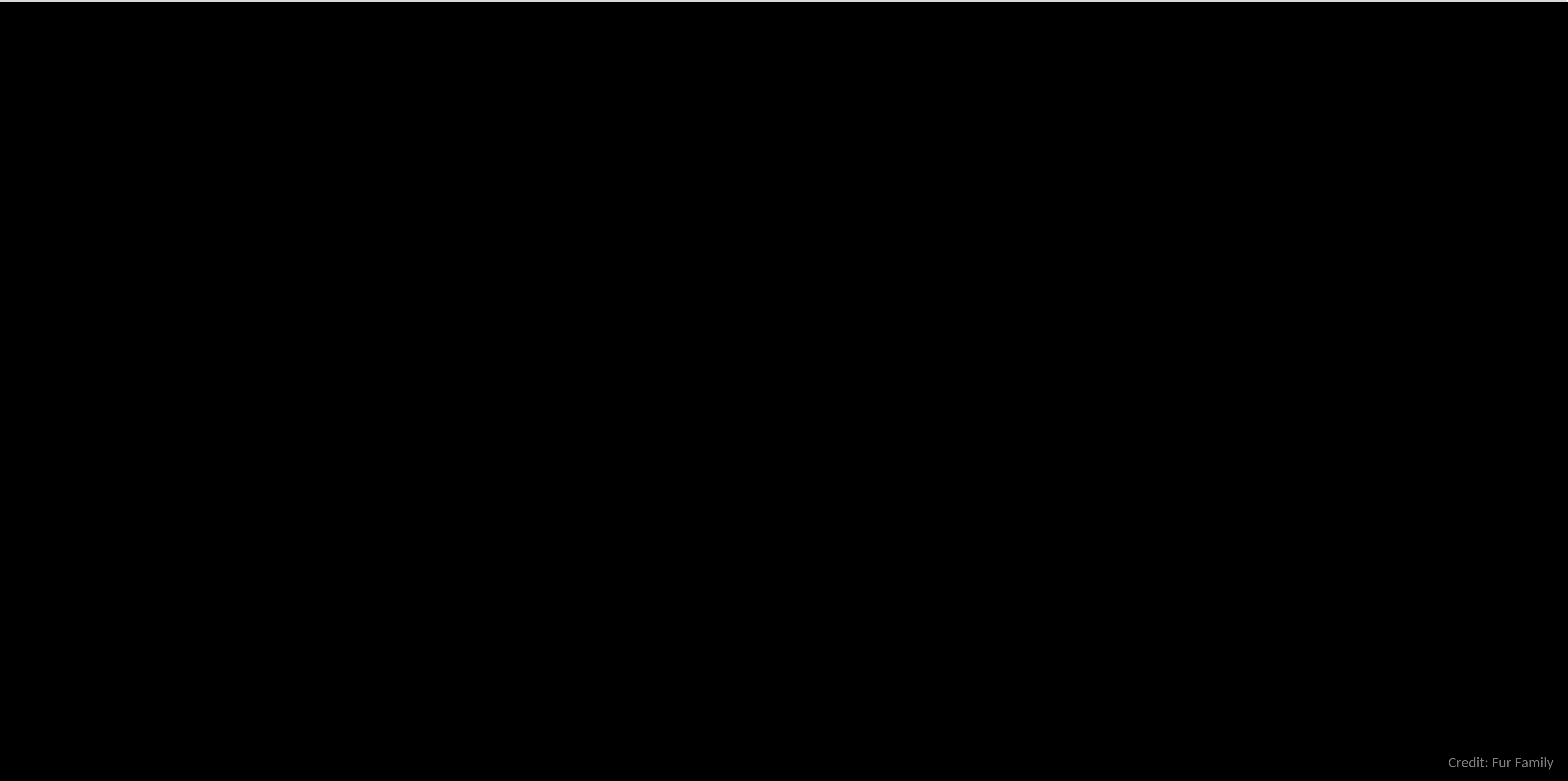
SCOTT GUAN

SCHOOL OF AEROSPACE ENGINEERING

GEORGIA INSTITUTE OF TECHNOLOGY

GAMEON! SEMINARS 2026

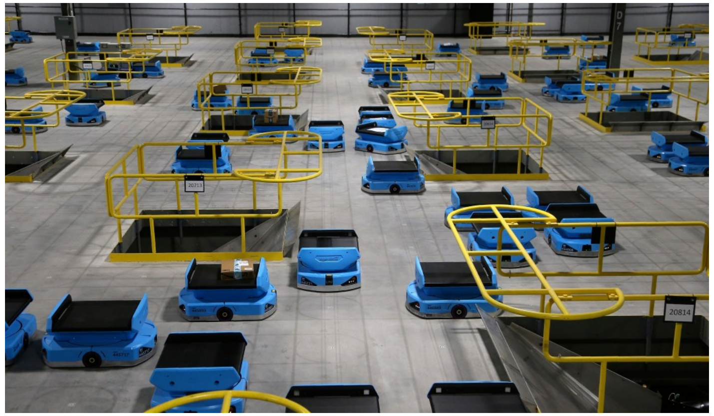
LARGE-POPULATION INTELLIGENCE … IS BEAUTIFUL



Credit: Fur Family

MULTI-AGENT SYSTEMS … ARE / WILL BE EVERYWHERE

Logistics



Warehouse automation

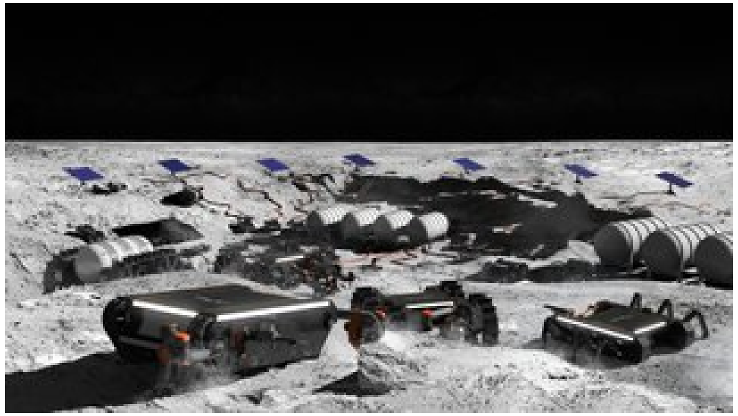
Transportation

Agriculture

Mining

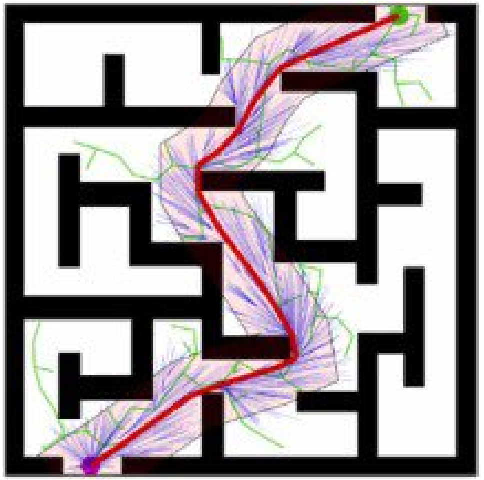
Surveillance

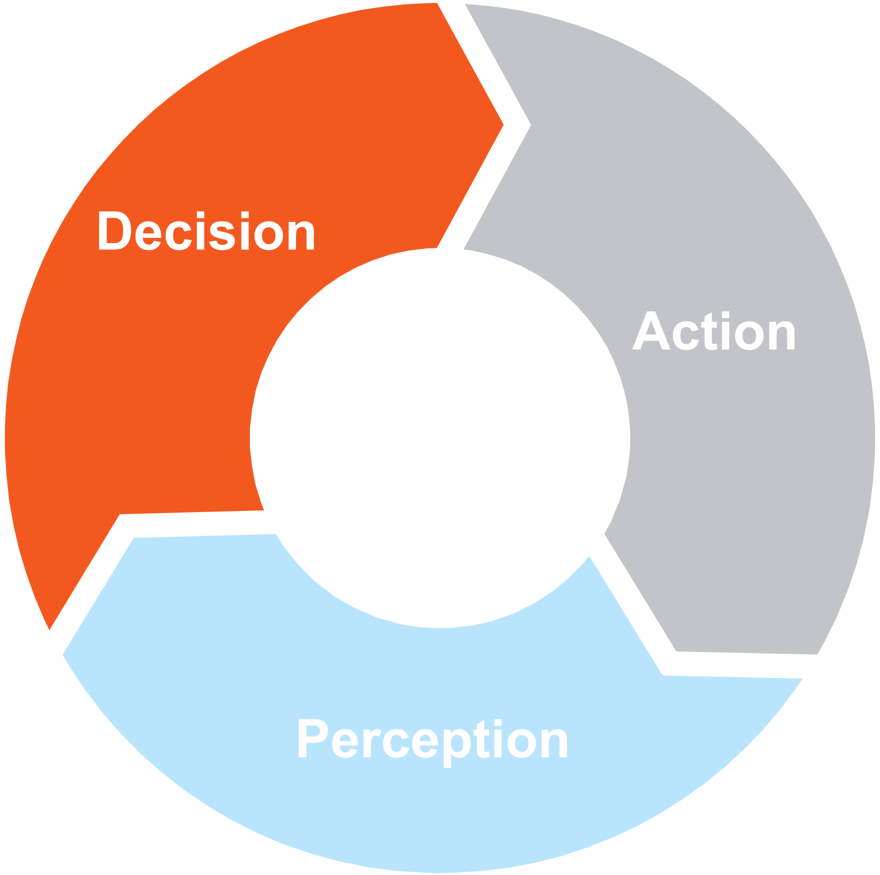
Military

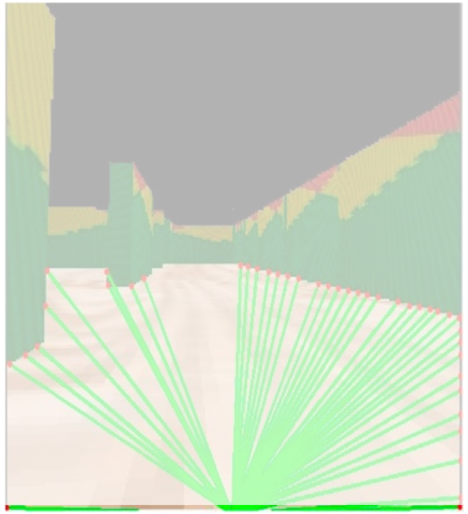


Space

PERCEPTION – DECISION – ACTION





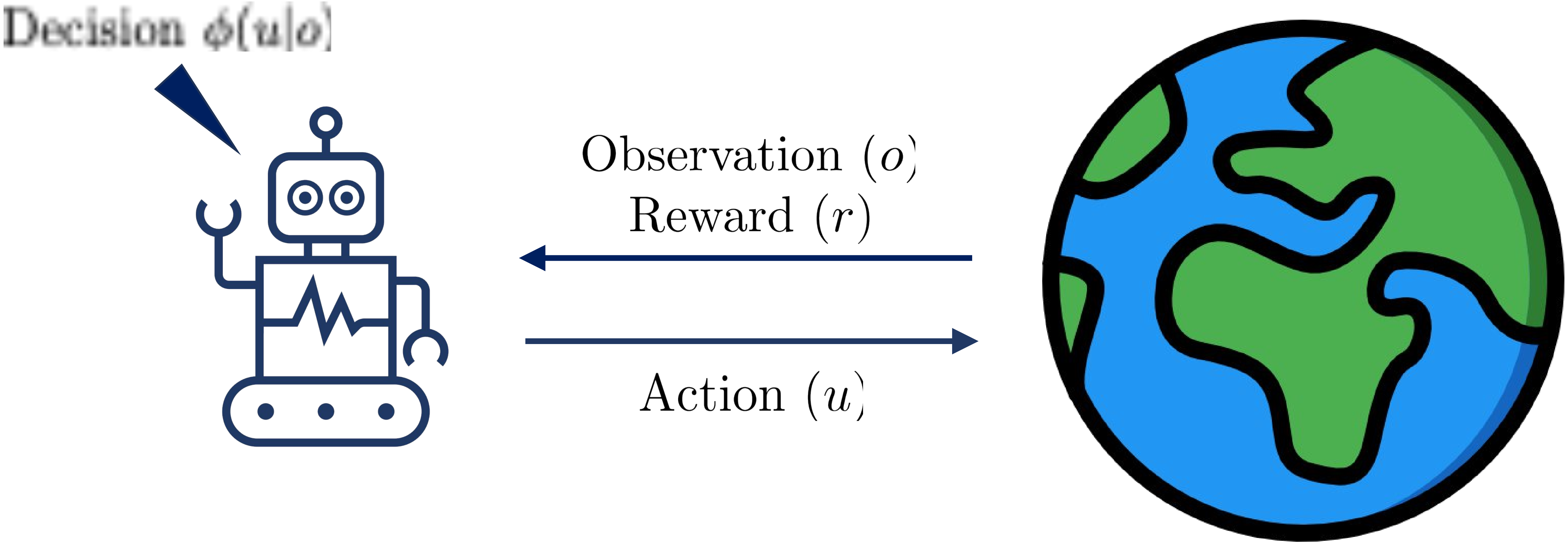


GROWING COMPLEXITY

Complex environment

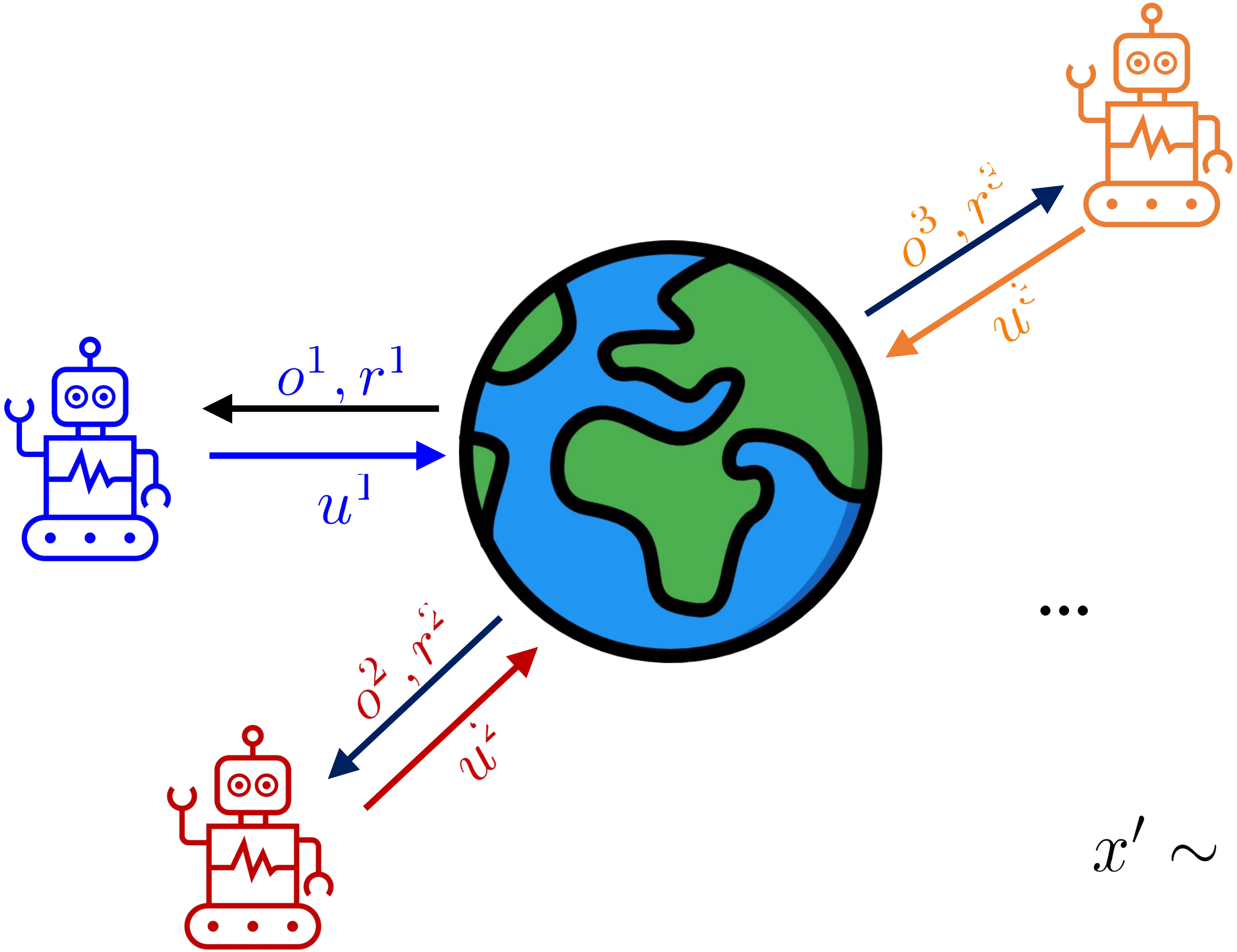
Large team size

MARKOV DECISION PROCESS (MDP)



$$x' \sim \mathcal{T}(x'|x, u)$$

FUNDAMENTAL CHALLENGE FOR MULTI-AGENT SYSTEMS

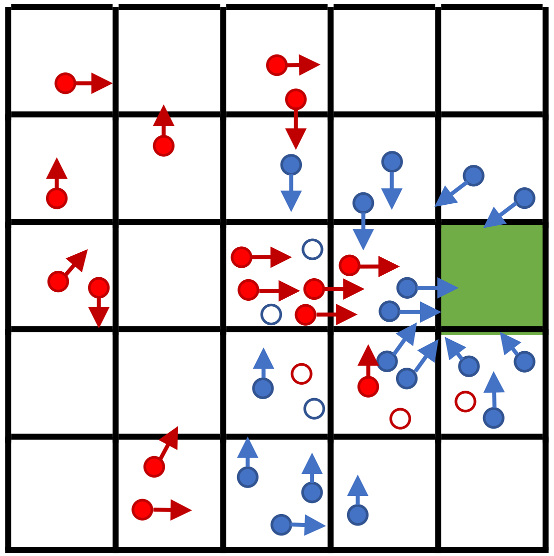


$$x' \sim \mathcal{T}(x' | x, u^1, u^2, u^3, \ldots)$$

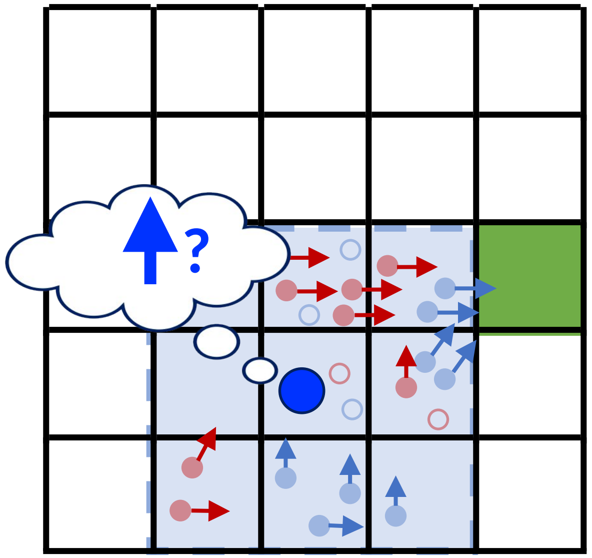
STOCHASTIC GAMES

- Game tuple $\langle \mathcal{X}^i, \mathcal{U}^i, \mathcal{R}^i, \mathcal{T}, H \rangle$
  - State & action spaces of agent i — $\mathcal{X}^i$ & $\mathcal{U}^i$
  - Reward for agent i — $\mathcal{R}^i : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$
  - Transition kernel — $\mathcal{T} : \mathcal{X} \times \mathcal{X} \times \mathcal{U} \to [0, 1]$
  - Finite horizon: — H

Joint state and action



- Game tuple $\langle \mathcal{X}^i, \mathcal{U}^i, \mathcal{R}^i, \mathcal{T}, H \rangle$
  - State & action spaces of agent i $\mathcal{X}^i$ & $\mathcal{U}^i$
  - Reward for agent i $\mathcal{R}^i : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$
  - Transition kernel $\mathcal{T} : \mathcal{X} \times \mathcal{X} \times \mathcal{U} \to [0, 1]$
  - Finite horizon: H
- Observation $O_t^i = \sigma^i(X_{0:t}^{1:N}, U_{0:t-1}^{1:N})$
- Strategy $\phi_t^i : \mathcal{U}^i \times \mathcal{O}_t^i \to [0, 1]$
- Optimization $\max_{\phi^i} \mathbb{E}\left[\sum_{t=0}^{H} \mathcal{R}^i(X_t, U_t)\right]$

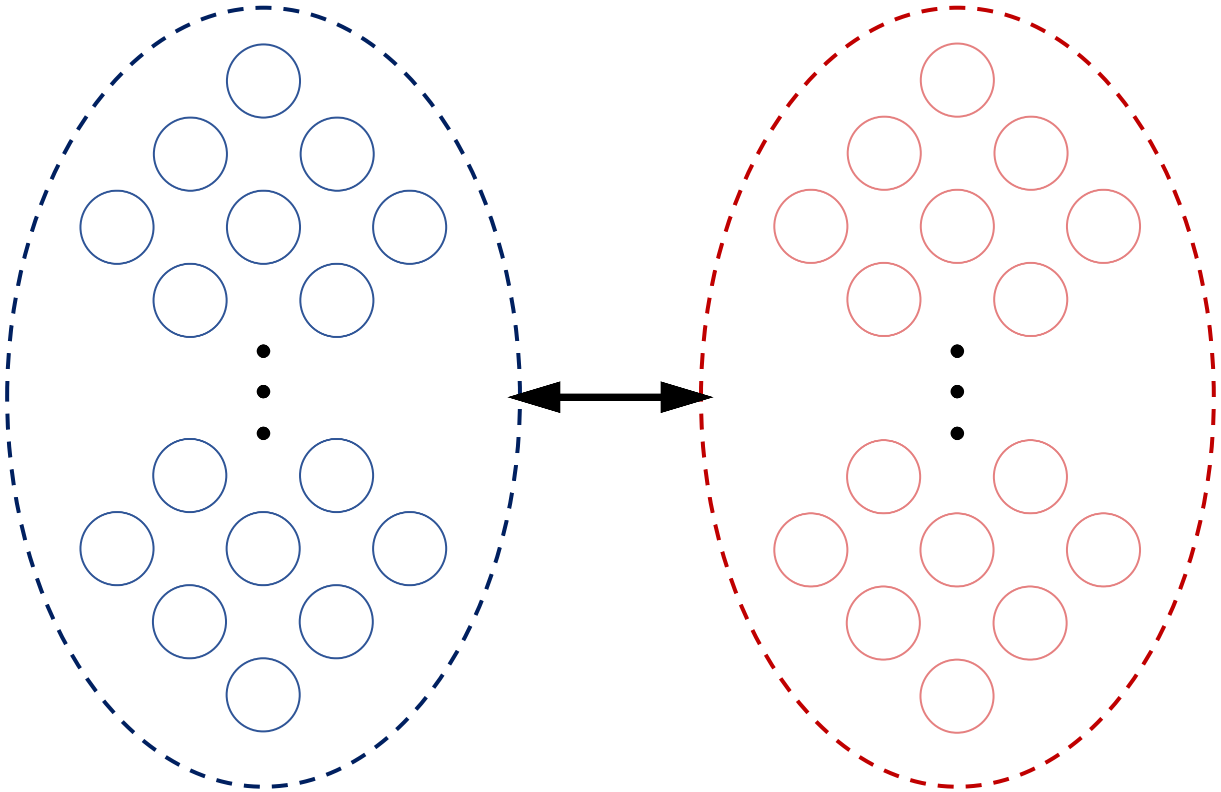


Challenges

- N coupled optimizations
- Complex information structure

In general, **NEXP!**

MEAN-FIELD TEAM GAMES



Zero-sum Mean-Field Team Games

- Large-population
- Generalization of zero-sum games
- Mixed collaborative-competitive setting

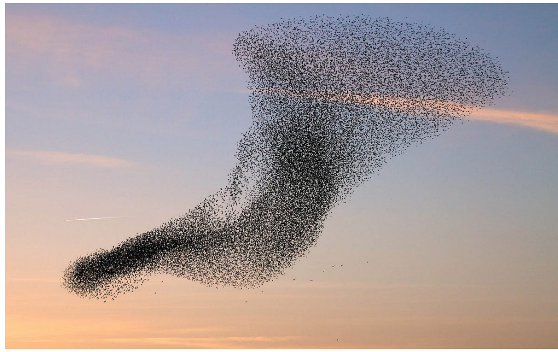
Overview of Results

- Identical Team Strategy Approximation
- Equivalent Two-Player Game
- ϵ-optimal performance

=> Scalable MARL algorithm: MF-MAPPO

MEAN-FIELD TEAM GAMES

- **Key Intuition:**
  Large population + homogeneous agents
  => Approximate system behaviors by agent distributions

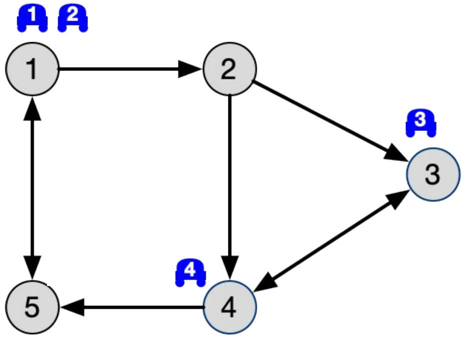


- **Mean-field:** fraction of agents on each state

$$\mu_t(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_x(x_t^i)$$

- **Weak coupling** through state distribution

$$r(x_t^i, \mu_t) \quad \& \quad f_t(x_{t+1}^i \mid x_t^i, u_t^i, \mu_t)$$



$$\mu_t = [\tfrac{1}{2}, 0, \tfrac{1}{4}, \tfrac{1}{4}, 0]$$

Mean-Field Literature

- Two major fields of research:
- MFG: Selfish agents

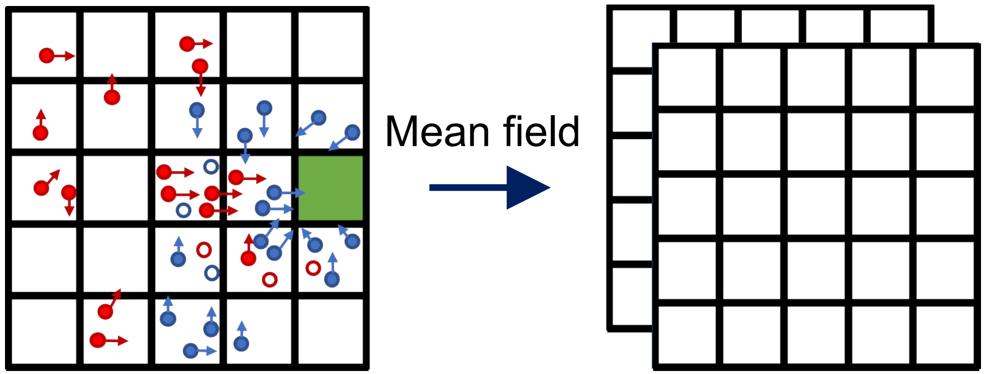
$$\max_{\pi^i} \mathbb{E}\left[\sum_{t=0}^{T_f} r(x_t^i, \mu_t)\right] \times N$$

- MFC/MFT: Collaborative agents

$$\max_{\{\pi^1,\ldots,\pi^N\}} \mathbb{E}\left[\sum_{t=0}^{T_f} r(\mu_t)\right]$$

- Mean-Field Equilibrium considers <u>single-agent</u> unilateral deviation

- Mixed collaborative and competitive setting:
  - Team level: competitive
  - Within each team: collaborative



- Optimization Objective

$$\underline{J}^{N*} = \max_{\phi^{N_1} \in \Phi^{N_1}} \min_{\psi^{N_2} \in \Psi^{N_2}} J^{N,\phi^{N_1},\psi^{N_2}}$$

team strategy

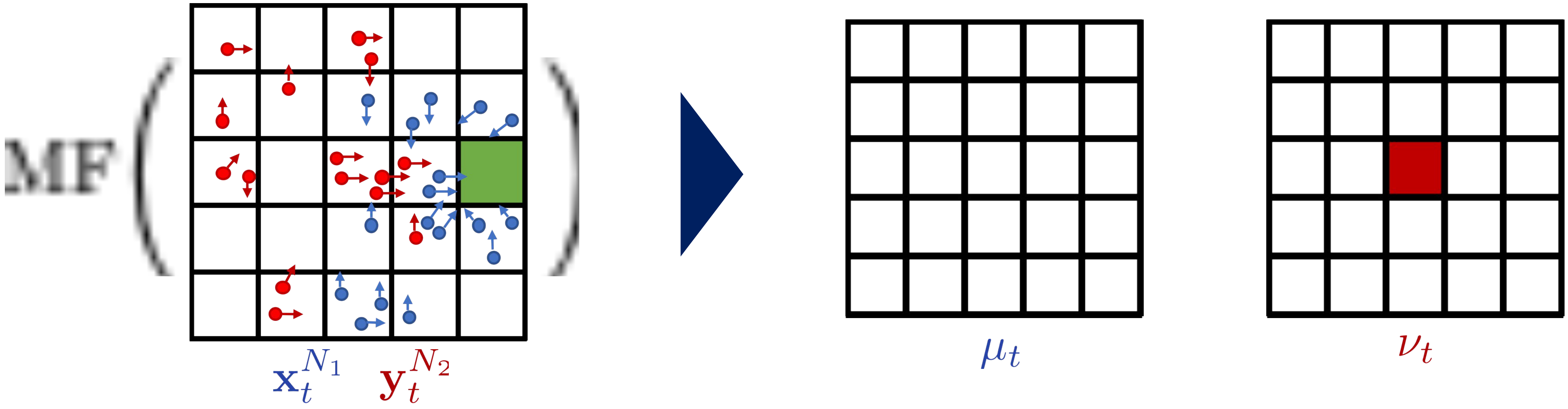
Key Challenges

- Team-level Deviation
- Non-identical strategies

> Complexity scales exponentially with the number of agents

MEAN-FIELD TEAM GAMES

- Zero-sum and simultaneous move
- Each team consists of N_j homogeneous agents
- Finite state and action spaces
- Finite horizon
- Weak coupling through the state distributions $f_t(x^i_{t+1} | x^i_t, u^i_t, \mu_t, \nu_t)$ & $\mathcal{R}(\mu_t, \nu_t)$



Mean-Field Sharing

Blue Agent Strategy:

Local Information

$$\phi_t^i(u|x_t^i, \mu_t, \nu_t)$$

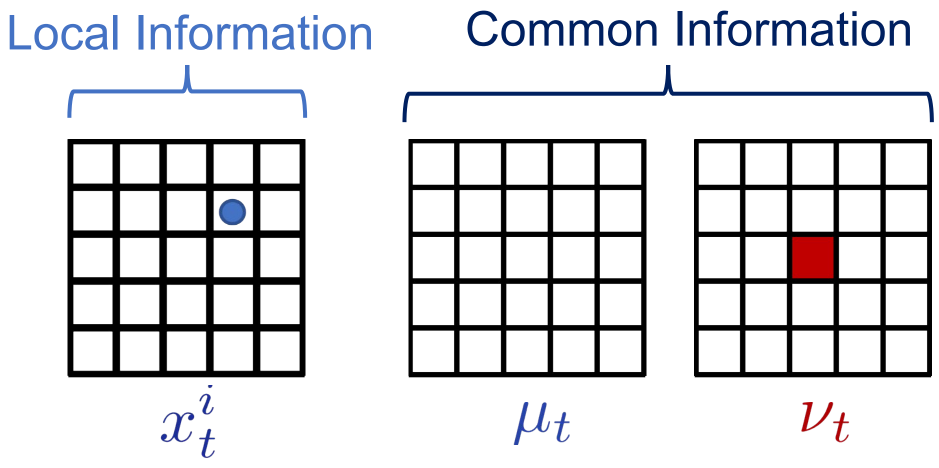
Common Information

Red Agent Strategy:

Common Information

$$\psi_t^j(v|y_t^j, \mu_t, \nu_t)$$

Local Information



Optimization Objective

$$\underline{J}^{N*} = \max_{\phi^{N_1} \in \Phi^{N_1}} \min_{\psi^{N_2} \in \Psi^{N_2}} J^{N, \phi^{N_1}, \psi^{N_2}}$$

blue team strategy

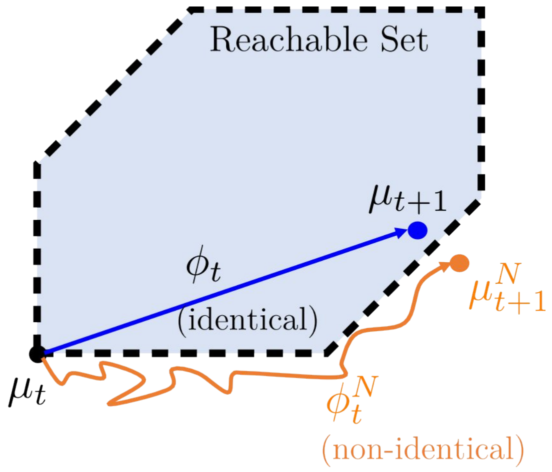
Challenge: Different agents (especially Opponents) can apply different strategies

HOW GOOD ARE IDENTICAL TEAM STRATEGIES?

Approximation Lemma

Given non-identical team strategy ϕ^{N_1}, there always exists an identical team strategy ϕ_t such that the distribution μ_{t+1} induced by ϕ_t is close to the distribution $\mu^{N_1}_{t+1}$ induced by $\phi^{N_1}_t$. In other words,

$$\inf_{\mu_{t+1}\in\mathcal{R}(\mu_t)} \mathbb{E}^{\phi_t^{N}}\left[d_{\mathrm{TV}}\left(\mu^{N_1}_{t+1},\mu_{t+1}\right)\right] \le \mathcal{O}\left(\sqrt{\frac{1}{N_1}}\right)$$

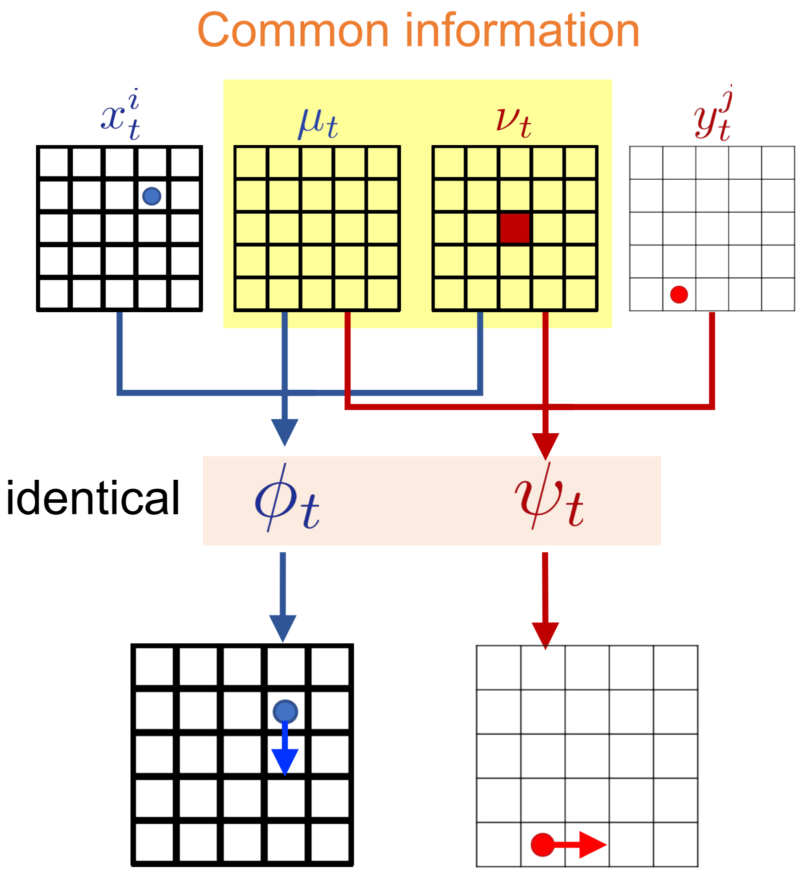


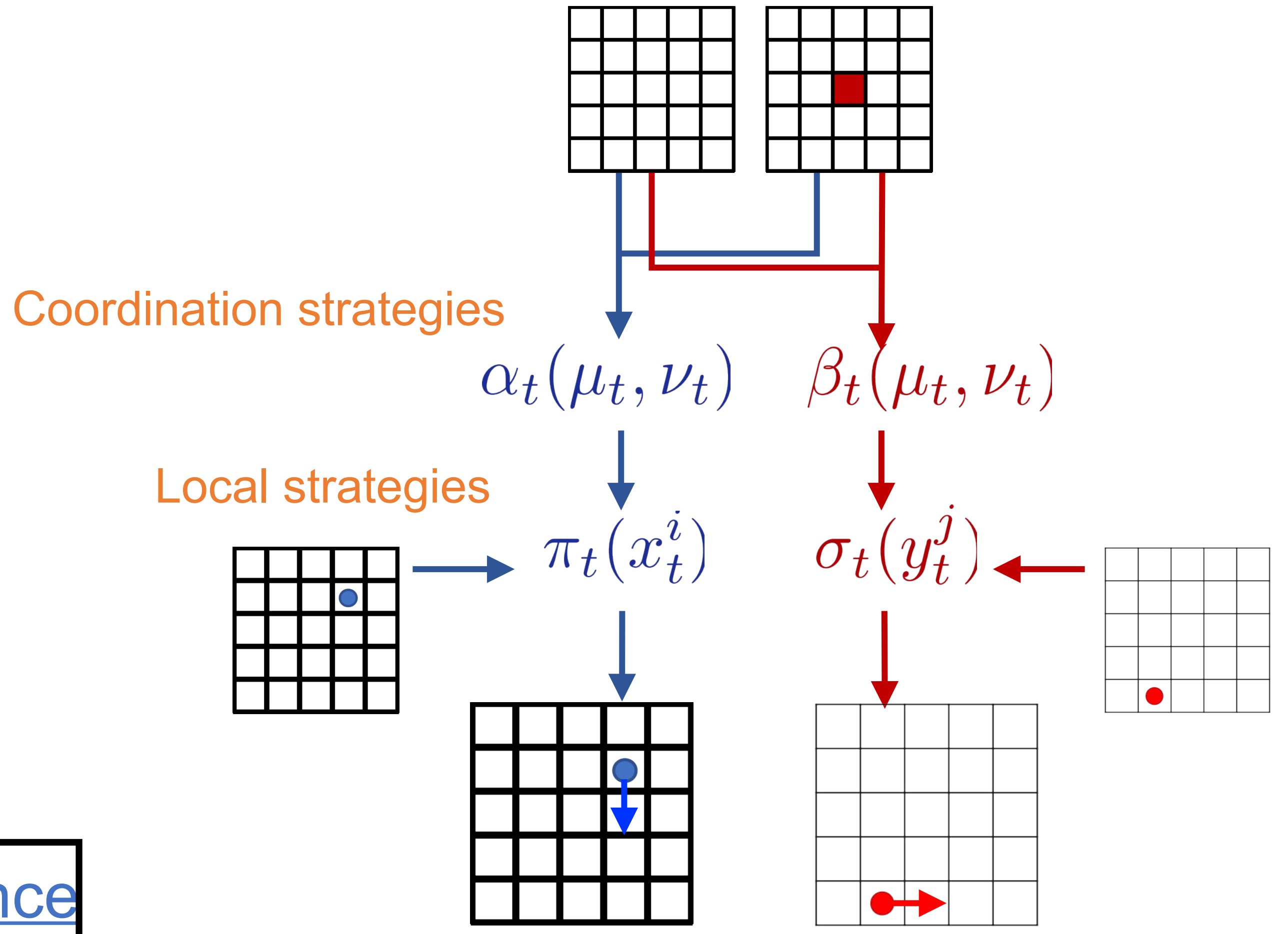
Key Takeaway

The set of identical team strategy is rich enough to approximate team distributions induced by non-identical team strategies when team size is large

Solve at the infinite-population limit with identical strategies

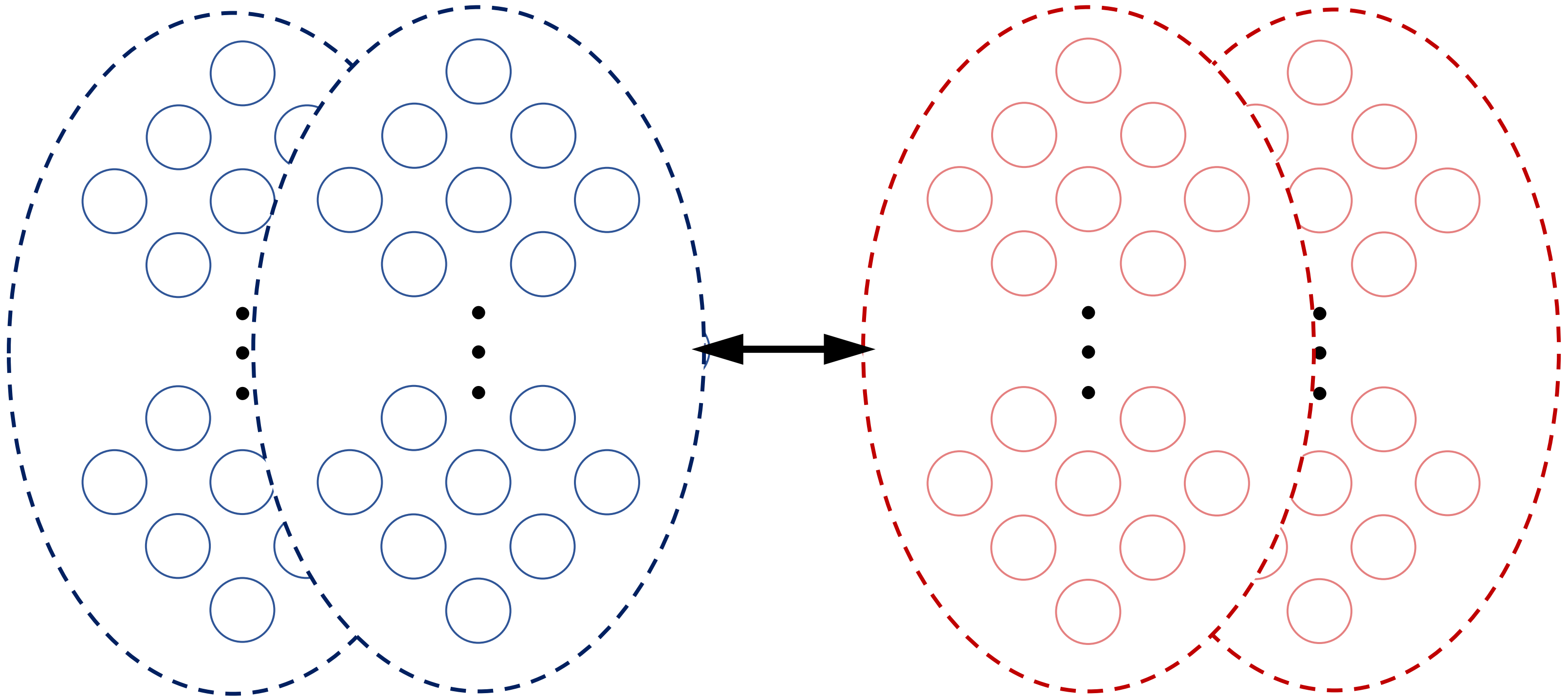
COMMON INFORMATION APPROACH





There is a one-to-one correspondence between ϕ_t and α_t

ZERO-SUM COORDINATOR GAME



- Can be solved via DP

PERFORMANCE GUARANTEES

Main Theorem

$$\underline{J}^{N*} = \max_{\phi^{N_1} \in \Phi^{N_1}} \min_{\psi^{N_2} \in \Psi^{N_2}} J^{N, \phi^{N_1}, \psi^{N_2}}$$

The optimal identical coordinator strategy α^* obtained from the infinite-population coordinator game induces an ϵ-optimal Blue team strategy in the finite-population game

Finite population

$$\underline{J}^{N*}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) \geq \min_{\psi^{N_2} \in \Psi^{N_2}} J^{N, \alpha^*, \psi^{N_2}}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) \geq \underline{J}^{N*}(\mathbf{x}^{N_1}, \mathbf{y}^{N_2}) - \mathcal{O}\left(\frac{1}{\sqrt{\underline{N}}}\right) \quad \forall\, \mathbf{x}^{N_1} \in \mathcal{X}^{N_1}, \mathbf{y}^{N_2} \in \mathcal{Y}^{N_2},$$

Red team uses non-identical policy to exploit

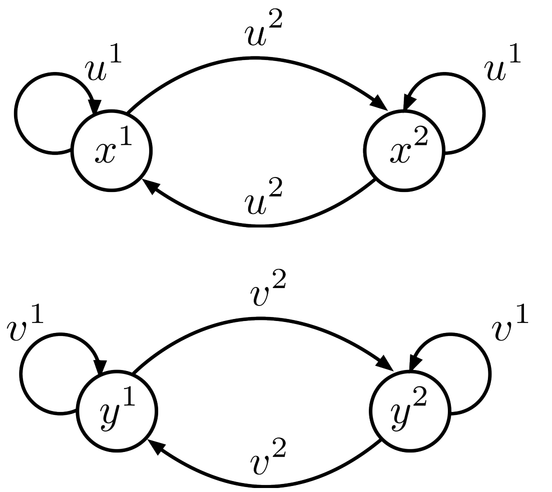
where $\underline{N} = \min\{N_1, N_2\}$

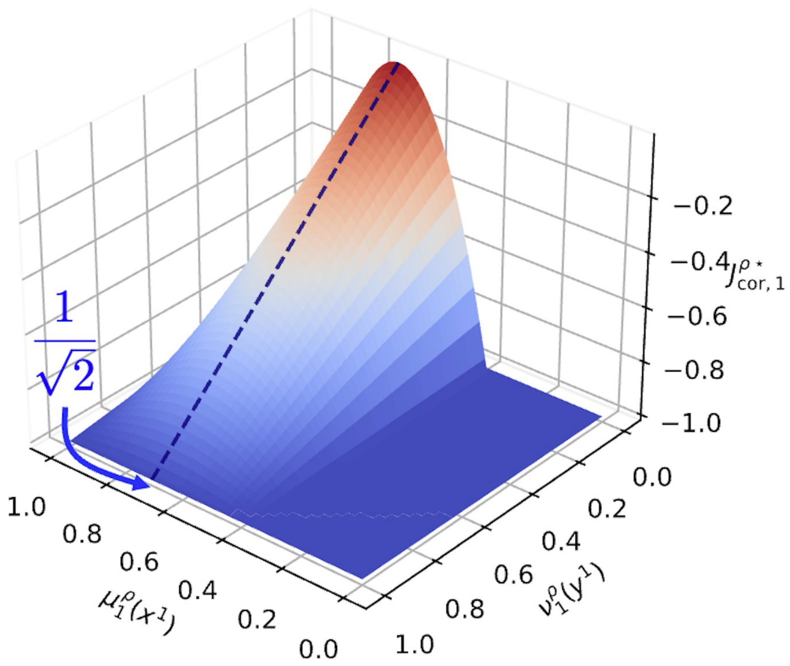
Key Takeaway

Mean-field team games can be solved considering only identical team strategies

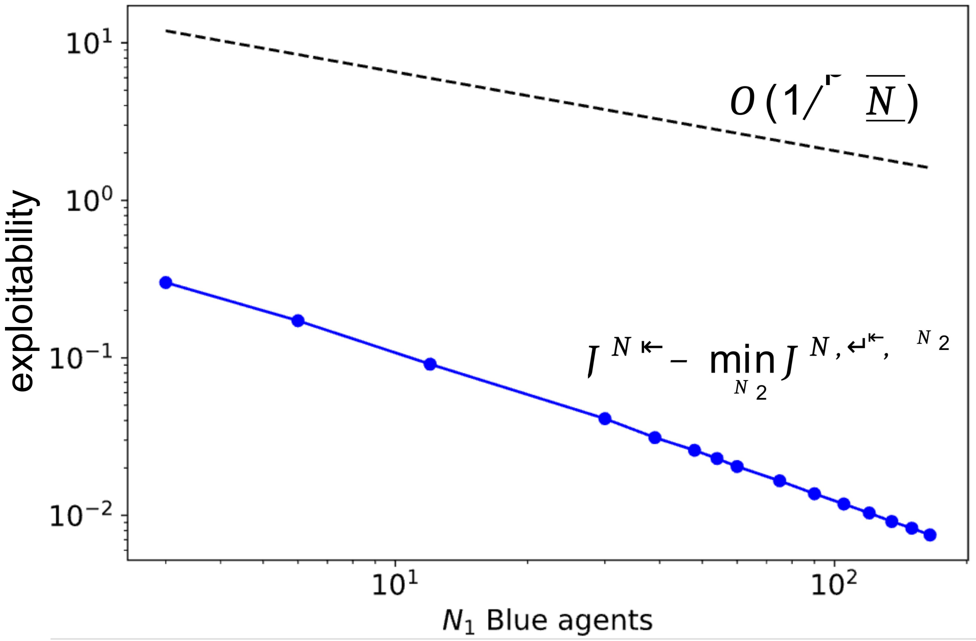
NUMERICAL EXAMPLE VIA DISCRETIZATION

- Two-node example
- Optimal Blue team strategy is to match distribution $[1/\sqrt{2}, 1-1/\sqrt{2}]$
  - Feasible only in infinite-population
  - Finite-population Blue optimal strategy is non-identical
- Exploitability is bounded by the predicted theoretical guarantee





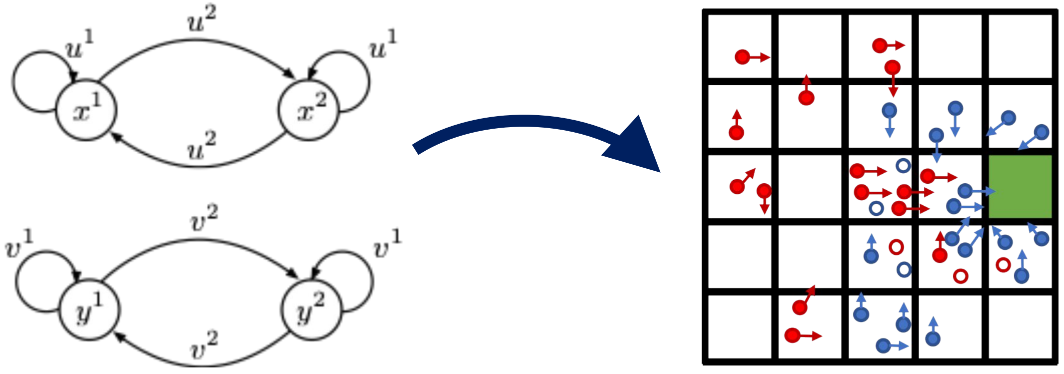
Coordinator game value



Exploitability vs. # of agents

MEAN-FIELD THEORY TO SCALE UP MARL

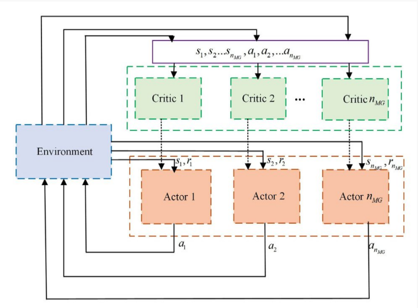
Realistic Large-scale Scenarios



Conventional Approach

Key limitations of MARL + CTDE

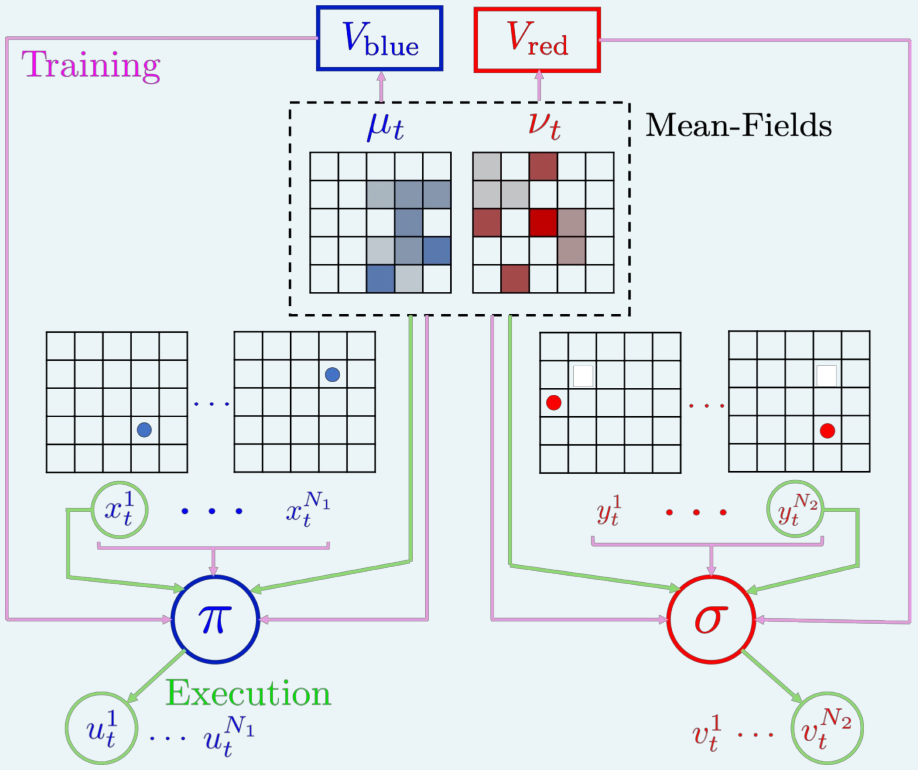
- Large amount of information to critic
- Individual agent has individual actor



MF-MAPPO

Mean-field Team-Game results:

- Value only depends on mean-field
- Identical team policy

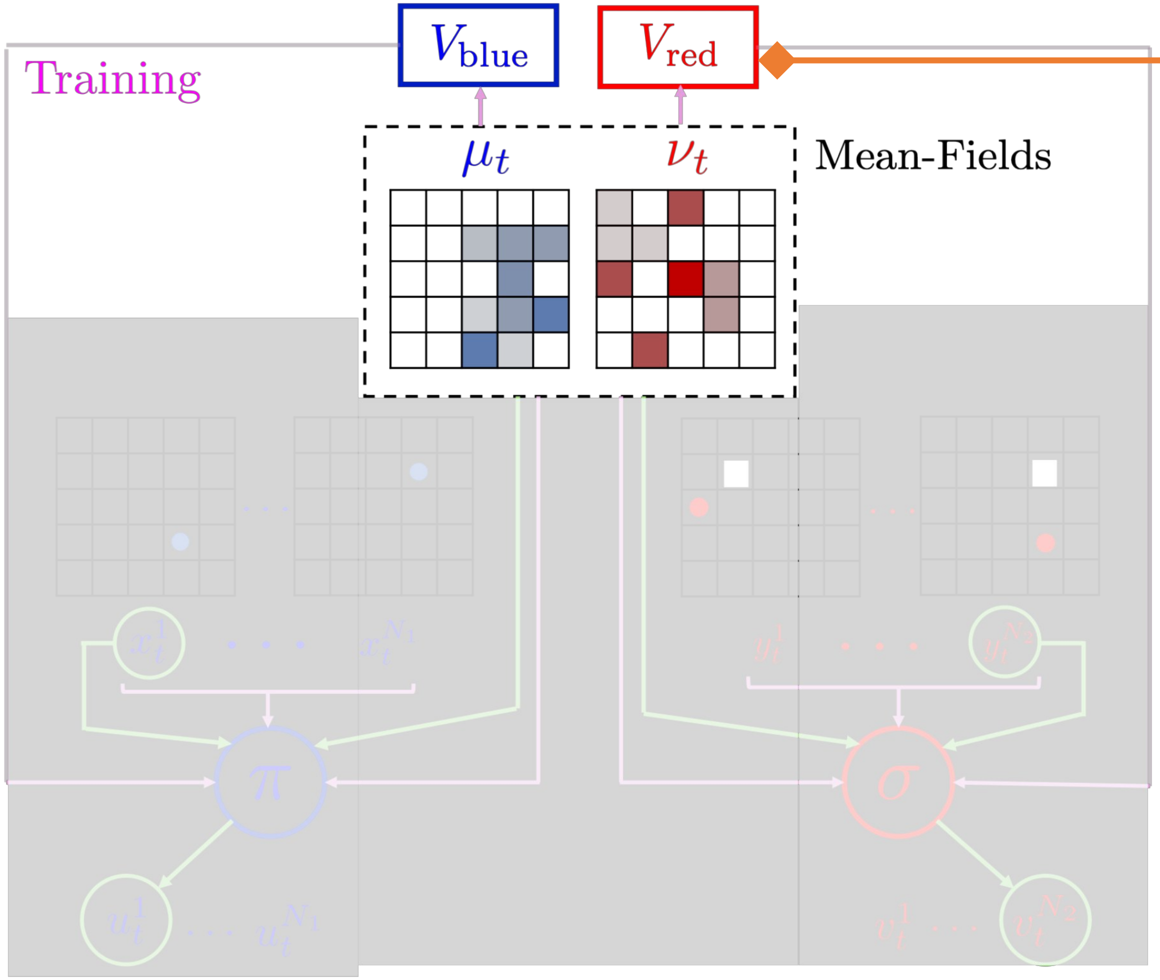


ARCHITECTURE OF MF-MAPPO: MINIMALLY-INFORMED CRITIC

MF-MAPPO

Mean-field Team-Game results:

- Value only depends on mean-field
- Identical team policy



Minimally-informed Critics

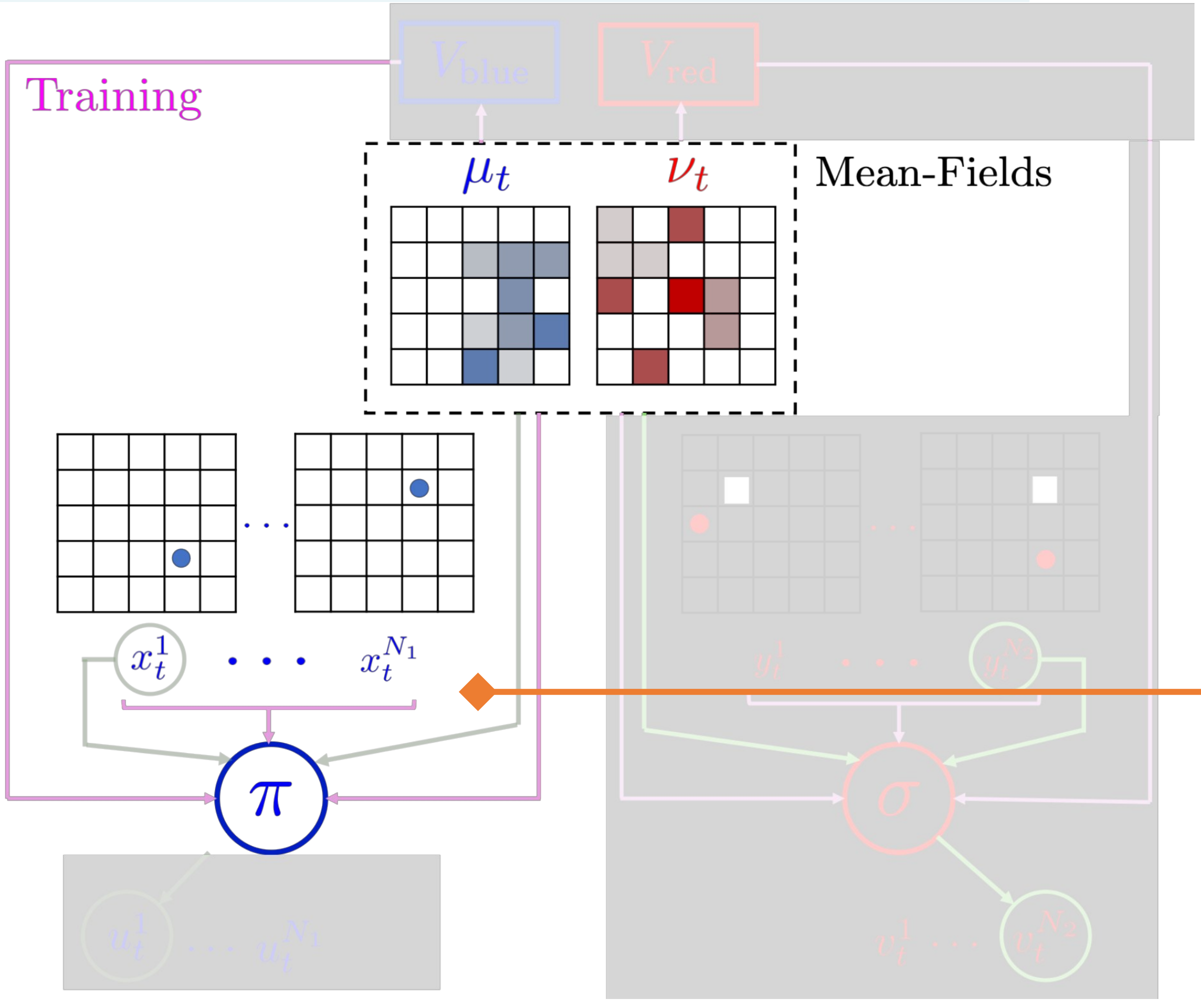
- PPO/value-based critic
- Only mean-fields as input
- Complexity does not scale with N_1, N_2

ARCHITECTURE OF MF-MAPPO: MINIMALLY-INFORMED CRITIC

MF-MAPPO

Mean-field Team-Game results:

- Value only depends on mean-field
- Identical team policy



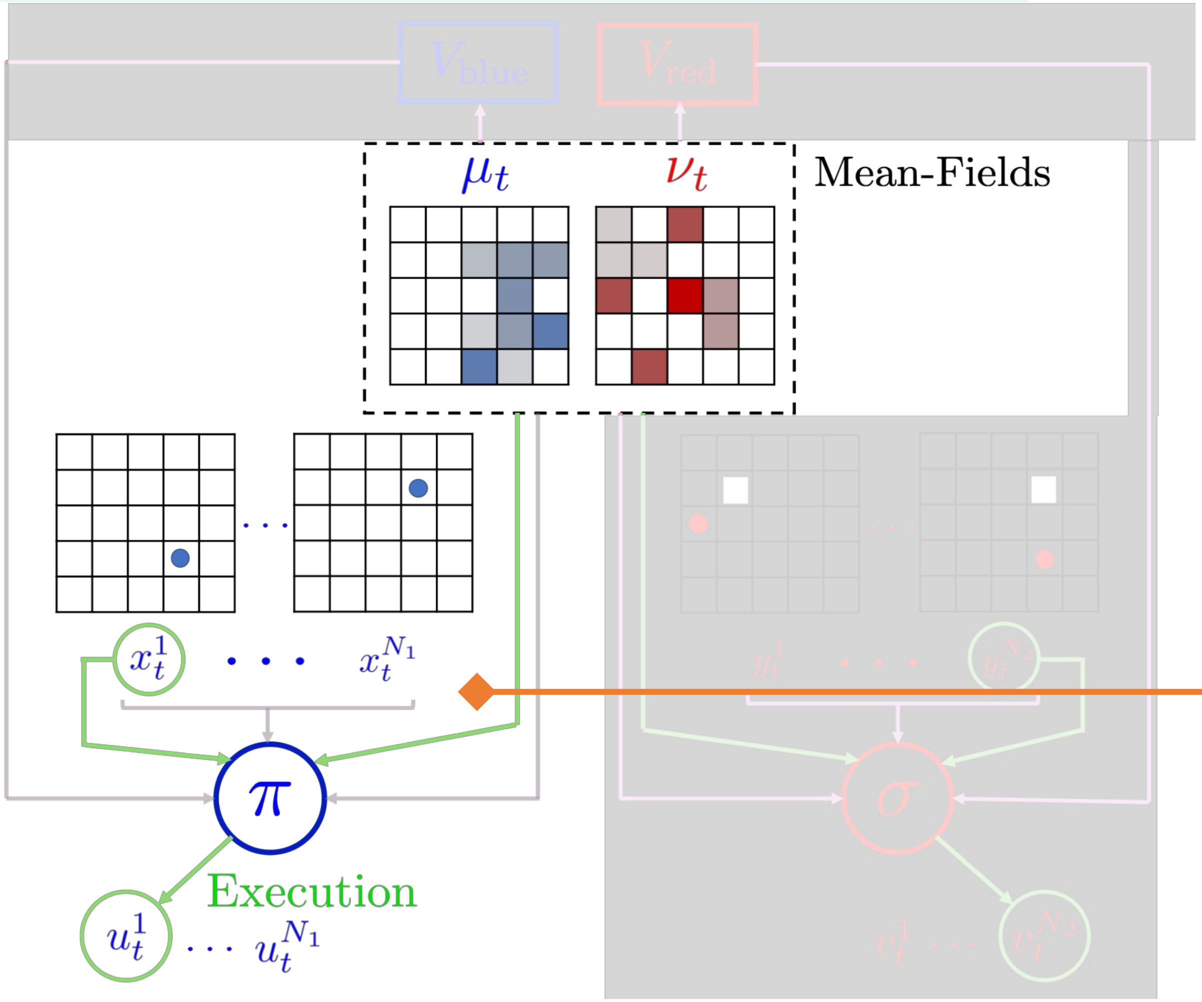
Shared Team-Actor

- Centralized training
- Simultaneous update

MF-MAPPO

Mean-field Team-Game results:

- Value only depends on mean-field
- Identical team policy



Shared Team-Actor

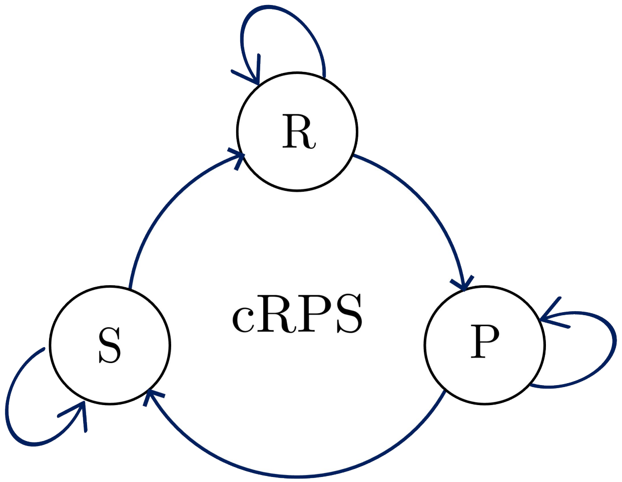
- Decentralized execution

VALIDATING SCENARIO: POPULATION RPS

Extends RPS to population game

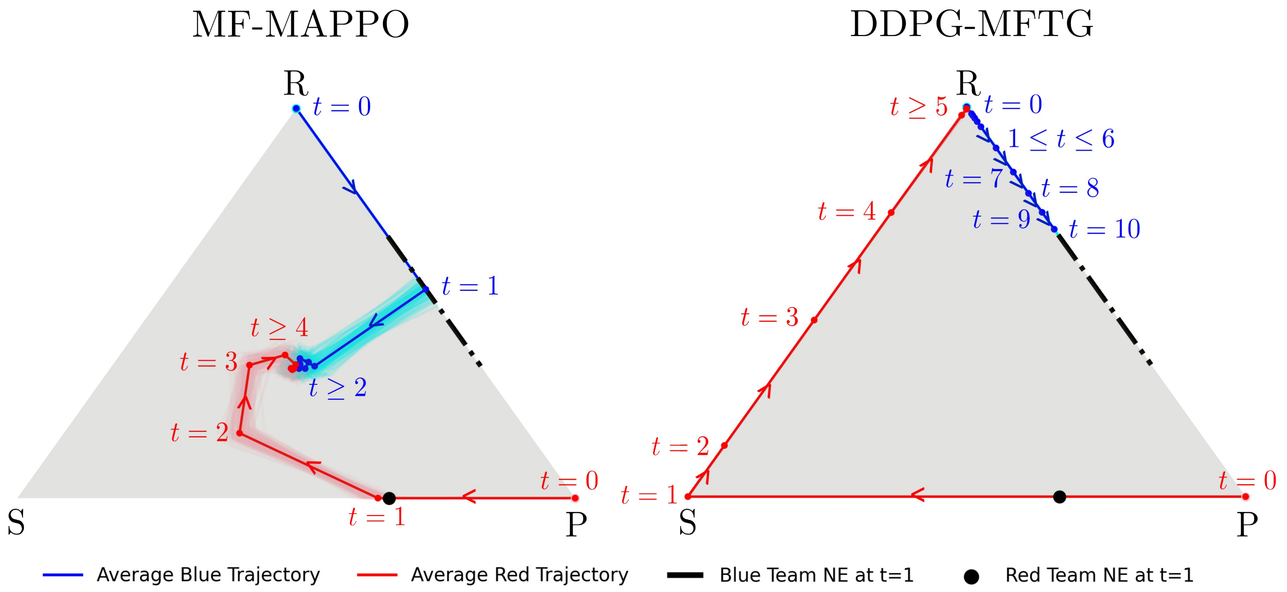
- Agent state space $\{\mathtt{R},\mathtt{P},\mathtt{S}\}$
- Action space $\{\mathtt{CW},\mathtt{Stay}\}$
- Running reward $r_t = \mu_t^\top A \nu_t$, where

$$A = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$$



Equilibrium Distribution

$$\mu^* = \nu^* = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$$



- MF-MAPPO converges to analytical equilibrium
- DDPG-based approach failed due to deterministic policies

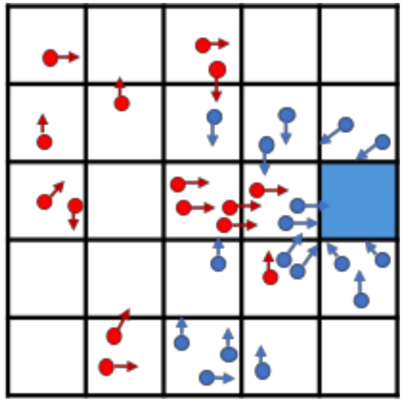
⇒ Mixed team policies are essential for convergence

BATTLE-FIELD SCENARIO WITH TARGET DEFENSE

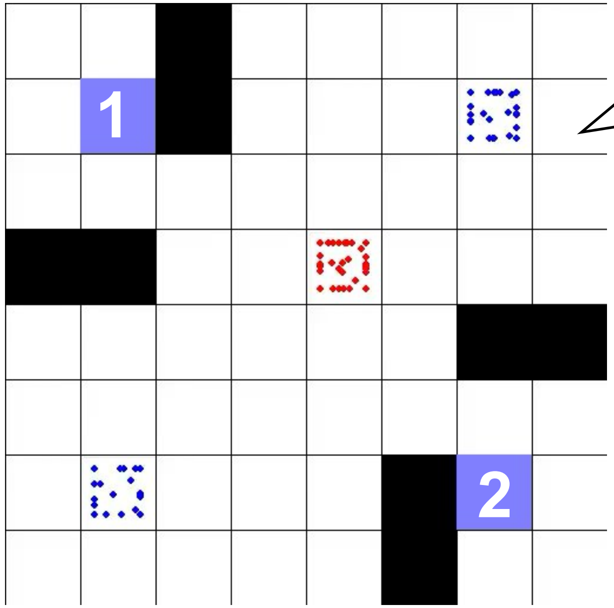
Game Setup

- Blue team reaches the target locations (blue)
- Red team defends the targets
- Agent gets deactivated based on numerical advantage at each cell

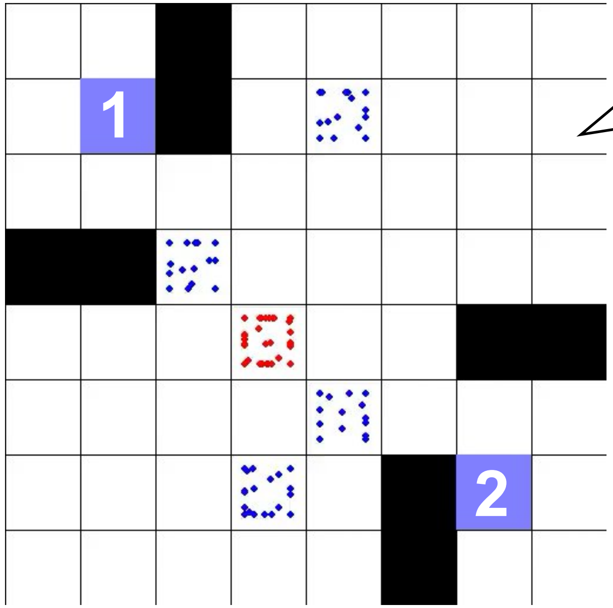
⇒ Teams must learn to avoid deactivation while attack/defend target



Learned Policy

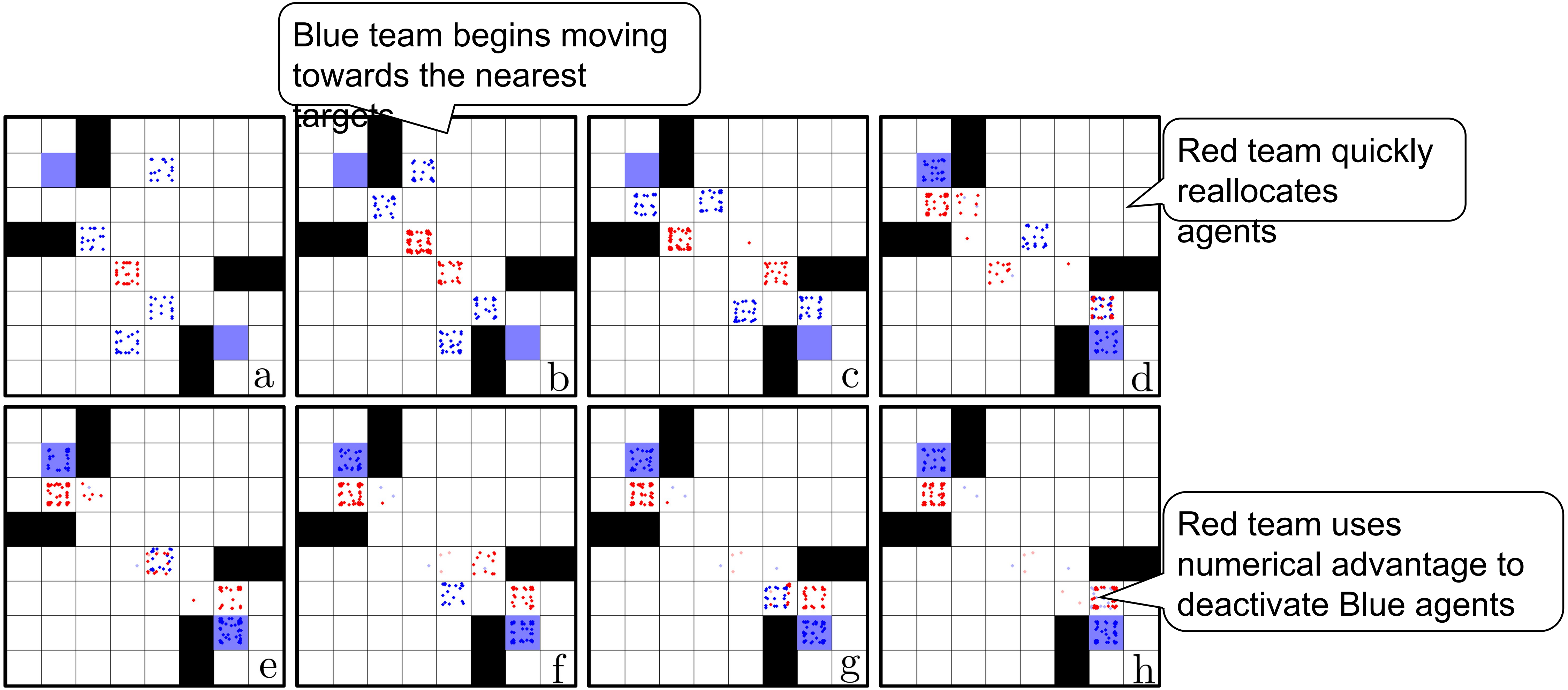


Blue team learns to fake an attack



Red team learns to block the corridors

HETEROGENEOUS BEHAVIORS FROM IDENTICAL TEAM POLICIES



Key Takeaway

- Mixed team policies ensure **convergence**
- Mixed identical team policies can still generate **heterogeneous behaviors**

CONCLUSION AND FUTURE WORK

Summary

- Mixed Collaborative-Competitive Team Game --- ZS-MFTG
- Mean-Field + Common information ⇒ Identical team strategy
- MF-MAPPO Algorithm
  - Minimally-informed critic using common information
  - Shared Team-Actor
  - Tractable learning for large-population teams

Future Work

- Limited/Partial observability
- Heterogeneous agents and sub-team roles/behaviors

ACKNOWLEDGEMENT

Panagiotis Tsiotras
Georgia Tech

Bhavini Jeloka
Georgia Tech

Mohammad Afshari
Georgia Tech

Ali Paknyat
U of Alabama

This work is supported by:

DEVCOM ARMY RESEARCH LABORATORY

Office of Naval Research Science & Technology

NSF

Q&A

Credit: Fur Family