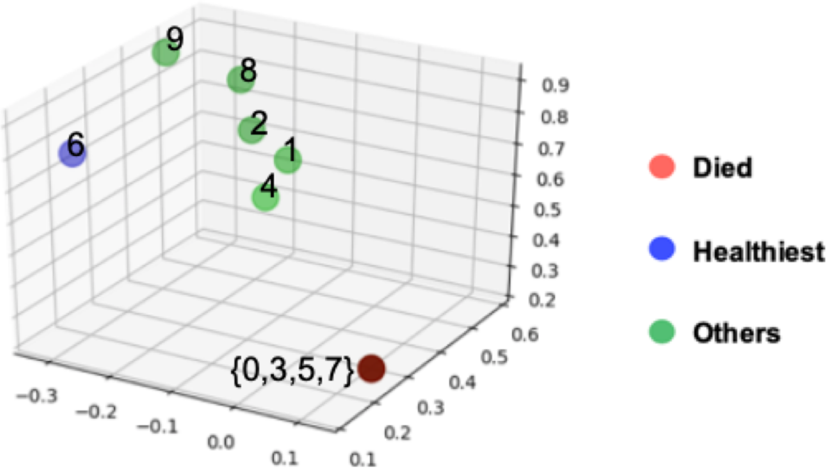
ROMA: Multi-Agent Reinforcement Learning with Emerging Roles

Tonghan Wang, Heng Dong, Victor Lesser, Chongjie Zhang

Tsinghua University, UMass Amherst



Multi-Agent Systems

Robot Football Game

Multi-Agent Assembly

One Major Challenge of Achieving Efficient MARL

- **Exponential blow-up of the state-action space**
  - The state-action space grows exponentially with the number of agents.
  - Learning a centralized strategy is not scalable.

- **Solution:**
  - Learning decentralized value functions or policies.

Decentralized Learning

- **Separate learning**
  - High learning complexity: Some agents are performing similar tasks from time to time;

- **Shared learning**
  - Share decentralized policies or value functions;
  - Adopted by most algorithms;
  - Can accelerate training.

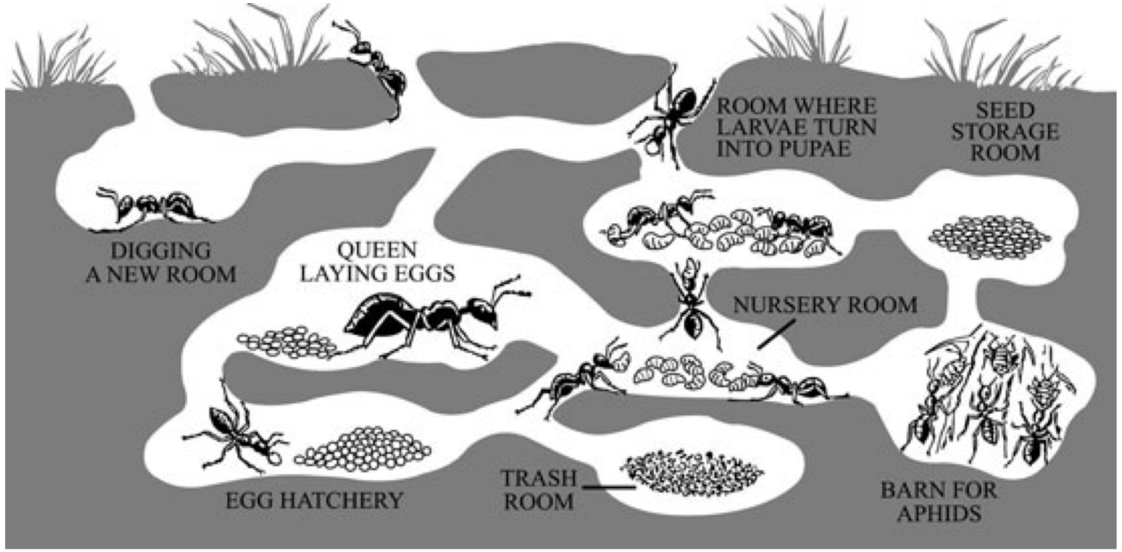
Drawbacks of Shared Learning

- **Parameter sharing**
  - Use a single policy to solve a task.
  - Inefficient in complex tasks. (Adam Smith's pin factory.)

- **An important direction of MARL**
  - Complex multi-agent cooperation needs ***sub-task specialization***.
  - Dynamic learning sharing among agents responsible for the same sub-task.

Draw Some Inspirations from Natural Systems

- **Ants**
  – Division of labor
- **Humans**
  – Share experience among people with the same vocation.



Role-Based Multi-Agent Systems

- **Previous work**
  - The complexity of agent design is reduced via task decomposition.
  - Predefine roles and associated responsibilities made up of a set of sub-tasks.

- **ROMA**
  - Incorporate role learning into multi-agent reinforcement learning.

Outline

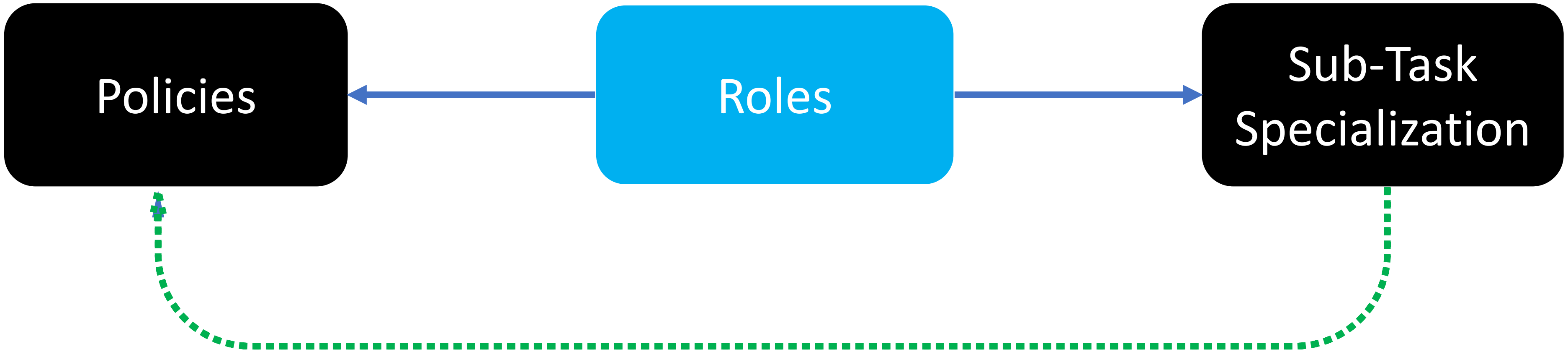
1. Motivation

2. Method

3. Results and Discussion

Our Idea

- Learn sub-task specialization.
- Let agents responsible for similar sub-tasks have similar policies and share their learning.
- Introduce roles.



Our method

- **Connection between roles and policies**
  - Generating role embeddings by a *role encoder* conditioned on local observations;
  - Conditioning agents' policies on individual roles.
- **Connection between roles and behaviors**
  - We propose two regularizers to enable roles to be:
    - Identifiable by behaviors
    - Specialized in certain sub-tasks

Identifiable Roles

- We propose a regularizer to maximize $I(\tau_i; \rho_i | o_i)$
- A lower bound:

$$I(\rho_i^t; \tau_i^{t-1} | o_i^t) \geq \mathbb{E}_{\rho_i^t, \tau_i^{t-1}, o_i^t}\left[\log \frac{q_\xi(\rho_i^t | \tau_i^{t-1}, o_i^t)}{p(\rho_i^t | o_i^t)}\right]$$

- In practice, we optimize

$$\mathcal{L}_I(\theta_\rho, \xi) = \mathbb{E}_{(\tau_i^{t-1}, o_i^t) \sim D}\left[\mathcal{CE}\left[p(\rho_i^t | o_i^t) \| q_\xi(\rho_i^t | \tau_i^{t-1}, o_i^t)\right] - H(\rho_i^t | o_i^t)\right]$$

Specialized Roles

- **We expect that, for any two agents,**
  - **Either** they have similar roles;
  - **Or** they have different behaviors, which are characterized by the local observation-action history.

- **However**
  - Which agents have similar roles?
  - How to measure the dissimilarity between agents' behaviors?

Specialized Roles

- **To solve this problem, we**
  - Introduce a learnable dissimilarity model d_ϕ
  - For each pair of agents, i and j, seek to maximize $I(\tau_j; \rho_i | o_j) + d_\phi(\tau_i, \tau_j)$
  - Seek to minimize $\|D_\phi\|_{2,0}$, the number of non-zero elements in $D_\phi = (d_{ij})$, where $d_{ij} = d_\phi(\tau_i, \tau_j)$

Specialized Roles

- Formally, we propose the following role embedding learning problem to encourage sub-task specialization:

$$\underset{\theta_\rho, \xi, \phi}{\text{minimize}} \quad \left\| D_\phi^t \right\|_{2,0}$$

$$\text{subject to} \quad I\left(\rho_i^t ; \tau_j^{t-1} \middle| o_i^t\right) + d_\phi\left(\tau_i^{t-1}, \tau_j^{t-1}\right) > U, \quad \forall i \neq j$$

- The specialization loss:

$$\mathcal{L}_D\left(\theta_\rho, \xi, \phi\right) = \mathbb{E}_{(\boldsymbol{\tau}^{t-1}, \boldsymbol{o}^t) \sim D, \boldsymbol{\rho}^t \sim \rho(\cdot|\boldsymbol{o}^t)}$$

$$\left[\left\| D_\phi^t \right\|_F - \sum_{i \neq j} \min\{q_\xi(\rho_i^t | \tau_j^{t-1}, o_i^t) + d_\phi\left(\tau_i^{t-1}, \tau_j^{t-1}\right), U\}\right]$$

Overall Optimization Objective

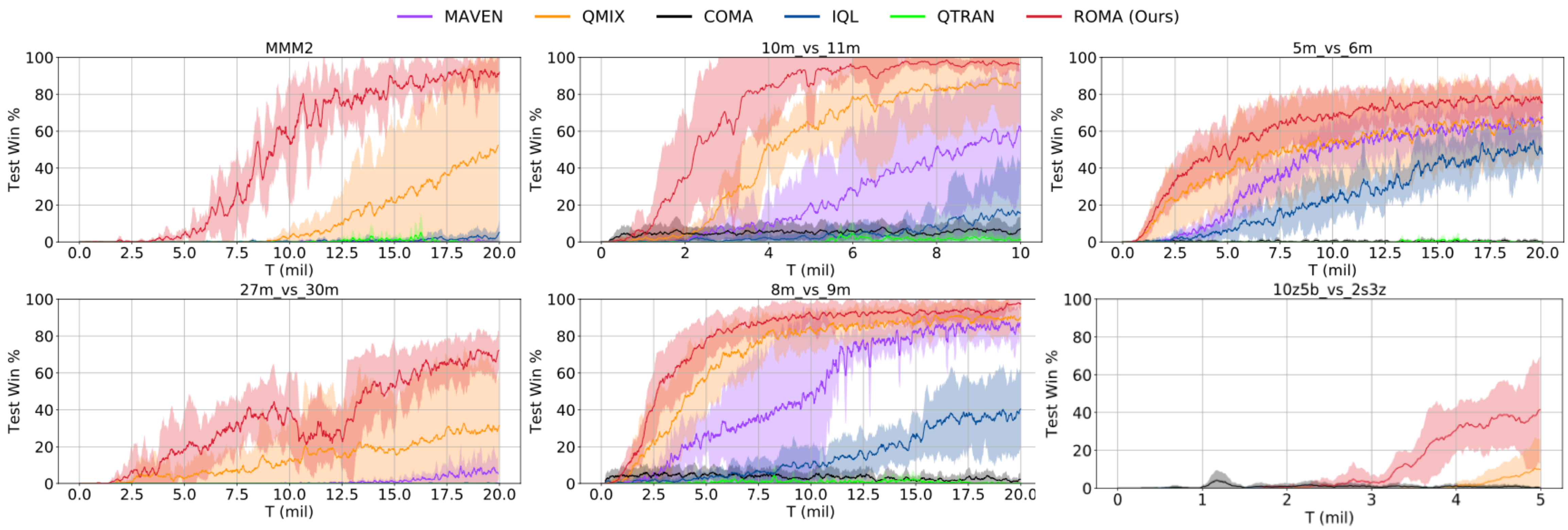
- **Overall Optimization Objective**
  - $\mathcal{L}(\theta) = \mathcal{L}_{TD}(\theta) + \lambda_I \mathcal{L}_I(\theta_\rho, \xi) + \lambda_D \mathcal{L}_D(\theta_\rho, \xi, \phi)$

Outline

1. Motivation
2. Methods
3. Results and Discussion

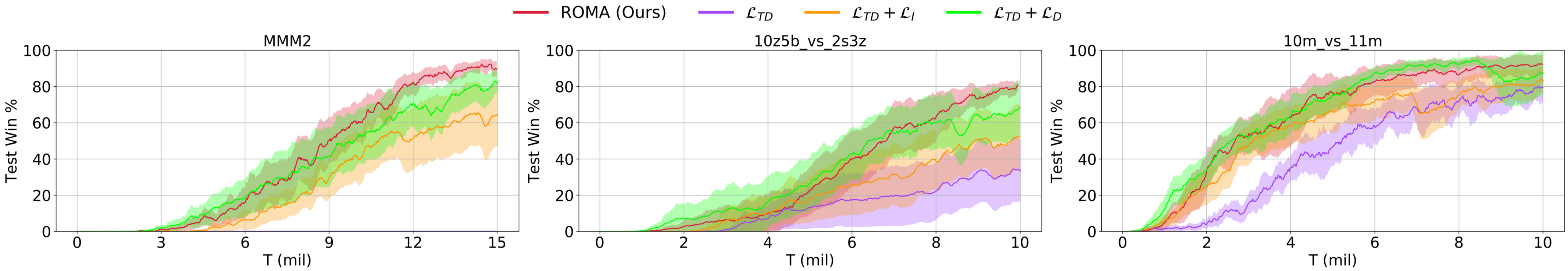
State-of-the-art performance on the SMAC benchmark

The SMAC Challenge



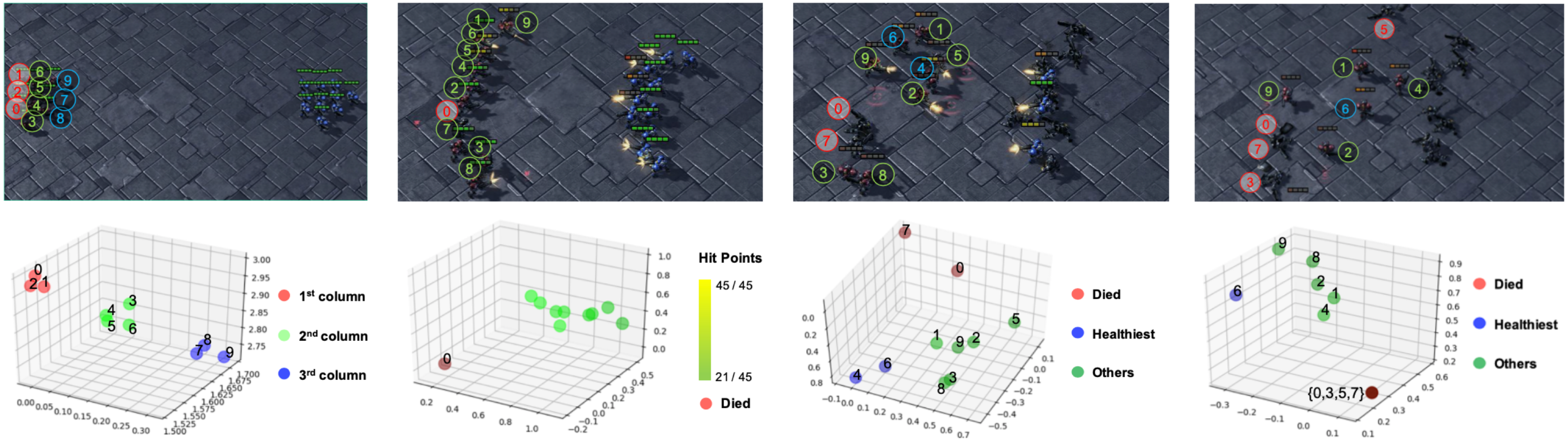
Ablation Study

Ablation Study



Role Representations

Dynamic Roles

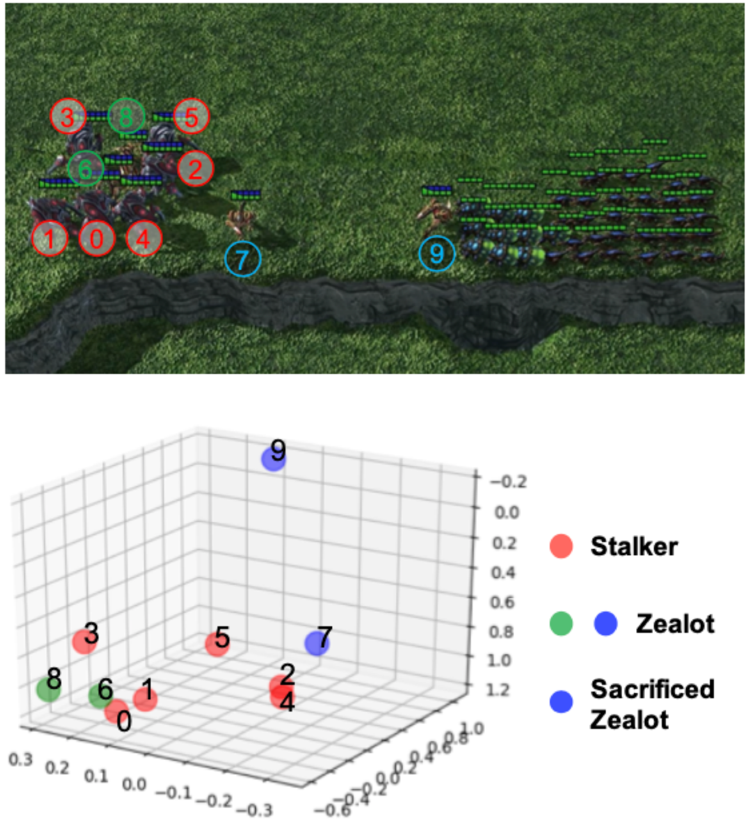


$t = 1$ $t = 8$ $t = 19$ $t = 27$

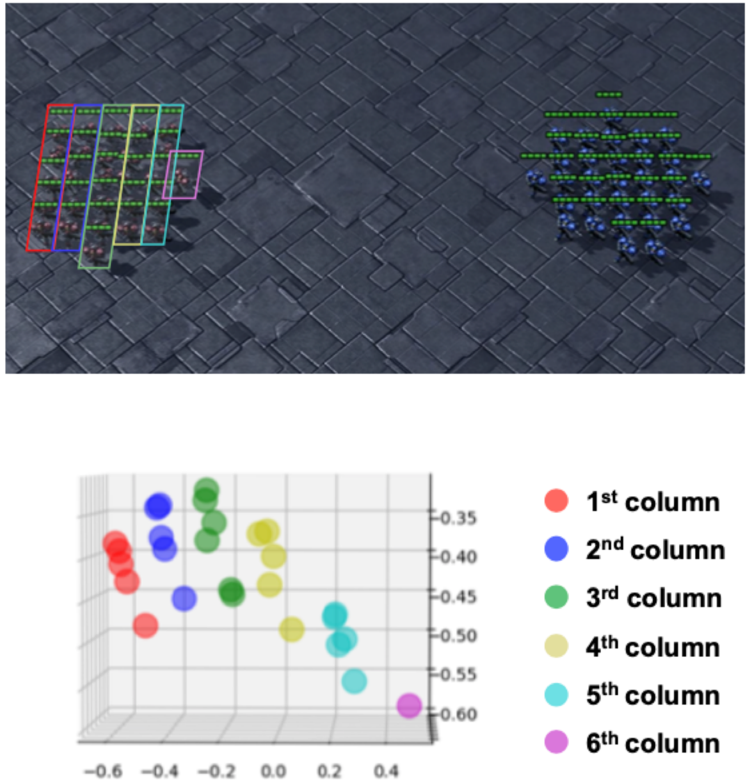
Specialized Roles

- **Learnable dissimilarity model:**
  - Map: MMM2;
  - Different unit types have different roles;
  - Learned dissimilarity between trajectories of *different* unit types: 0.9556 ± 0.0009;
  - Learned dissimilarity between trajectories of *the same* unit type: 0.0780 ± 0.0019.

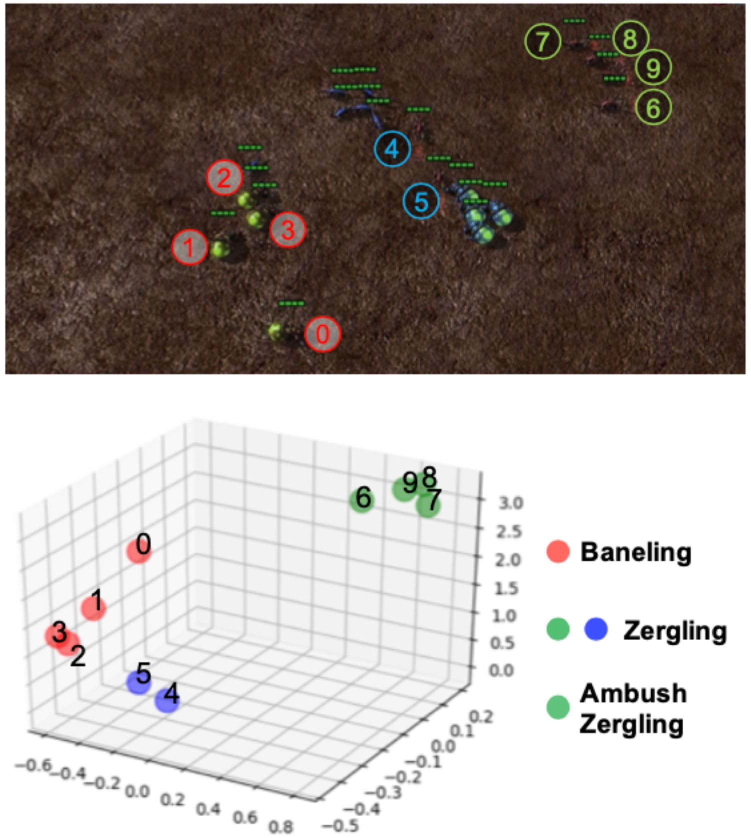
Specialized Roles



(a) Strategy: sacrificing Zealots 9 and 7 to minimize Banelings' splash damage.



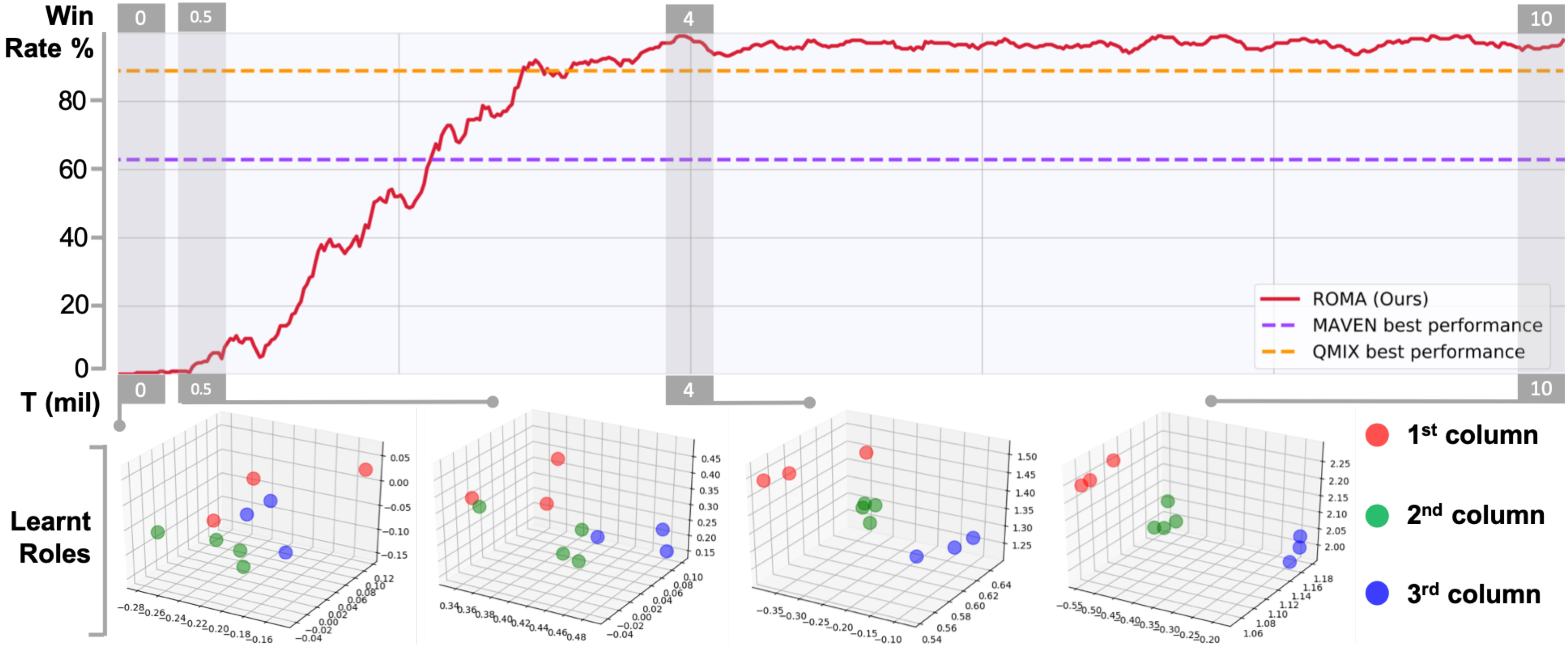
(b) Strategy: forming an offensive concave arc quickly



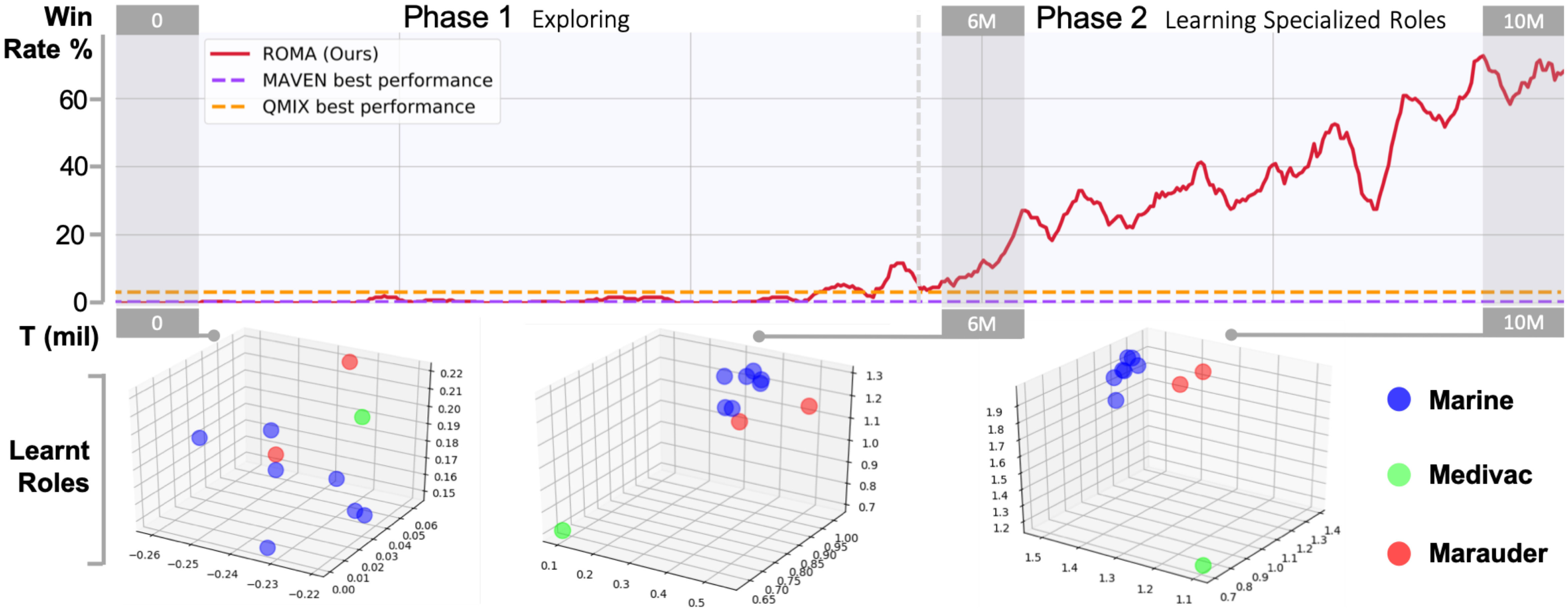
(c) Strategy: green Zerglings hide away and Banelings kill most enemies by explosion.

Multi-Agent Reinforcement Learning with Emerging Roles

Role Emergence



Role Emergence



Game Replays

27m_vs_30m (27 Marines vs. 30 Marines)

For more experimental results.

Welcome to our website:

- https://sites.google.com/view/romarl