Multi-agent Coordination via Flow Matching

Dongsu Lee[1] Daehee Lee[2] Amy Zhang[1]

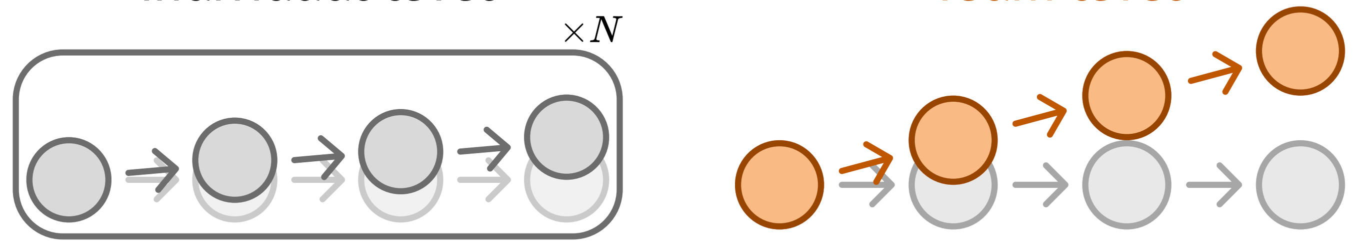
[1]UT Austin [2]Sungkyunkwan Univ.

Blog Code

Data-driven multi-agent RL

Problem definition for "offline MARL and offline-to-online MARL"

Objective: extract a set of policies to maximize team reward while staying close to data distribution



Challenge: accumulation of small individual deviations can cause critical out-of-distribution for the entire team

Prior connection

Tradeoff between coordinating performance and cost

The simplest method: extension of offline RL

$$\pi_{\theta_i}(a_i|o_i) = \mathcal{N}(a_i; \mu_{\theta_i}(o_i), \sigma_{\theta_i}(s_i))$$

Individual Gaussian policy cannot capture multi-modality

low expressivity :(

The recent method: diffusion modeling

$$\pi_\theta(\mathbf{a}^{0:K}|\mathbf{o}) = p(\mathbf{a}^K)\prod_{k=1}^{K} p_\theta(\mathbf{a}^{k-1}|\mathbf{a}^k,\mathbf{o})$$

$$p_\theta(\mathbf{a}^{k-1}|\mathbf{a}^k,\mathbf{o}) = \mathcal{N}(\mathbf{a}^{k-1}; \boldsymbol{\mu}_\theta(\mathbf{a}^k,k,\mathbf{o}), \boldsymbol{\Sigma}_\theta)$$

Iterative denoising steps are computationally expensive

low efficiency :(

Can't we adopt flow matching?

$$\mathcal{L}_{\text{Flow-BC}}(\phi) = \mathbb{E}\left[\|v_\phi(t,\mathbf{o},\mathbf{x}^t) - (\mathbf{a}-\mathbf{x}^0)\|_2^2\right]$$

Joint flow breaks decentralized manner

$$\mathcal{L}_{\text{Flow-BC}}(\phi_i) = \mathbb{E}\left[\|v_{\phi_i}(t,o_i,x^t) - (a_i - x^0)\|_2^2\right]$$

Individual flow can't consider coordination

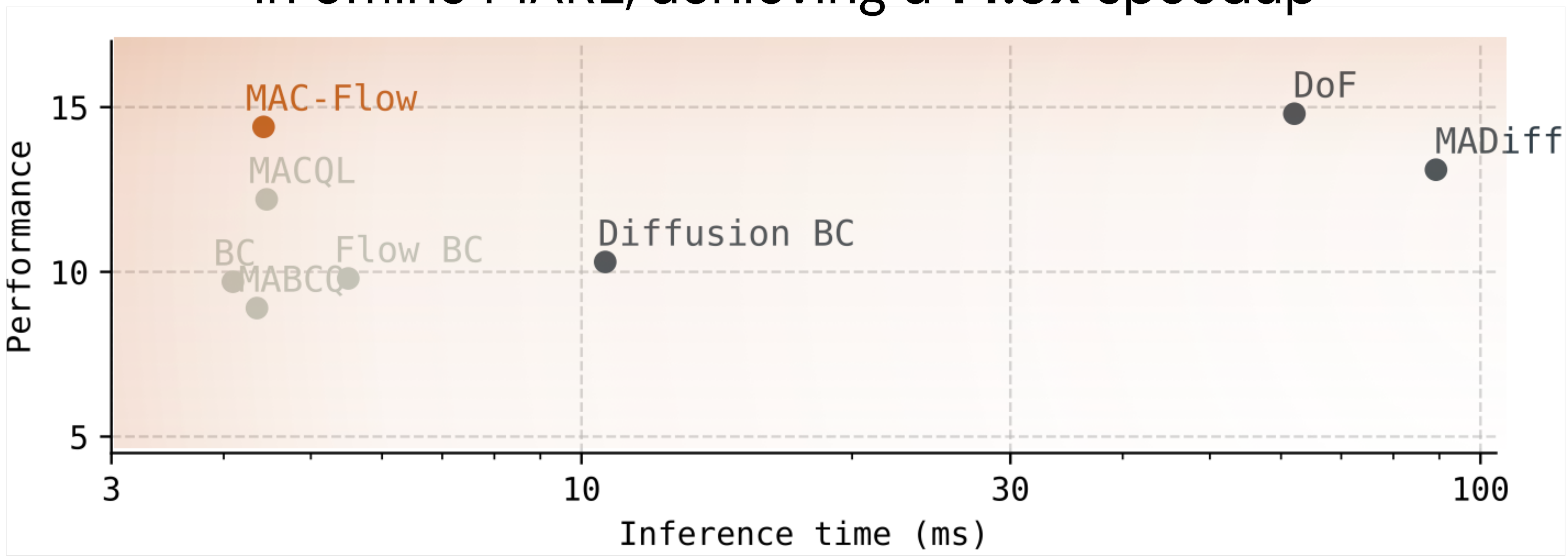
$$\mathcal{L}(\pi_i) = \mathbb{E}[-Q_{\text{tot}}(s,\mathbf{a})] + \alpha\mathcal{L}_{\text{Flow-BC}}(\phi_i)$$

BPTT with iterative sampling is unstable and inefficient

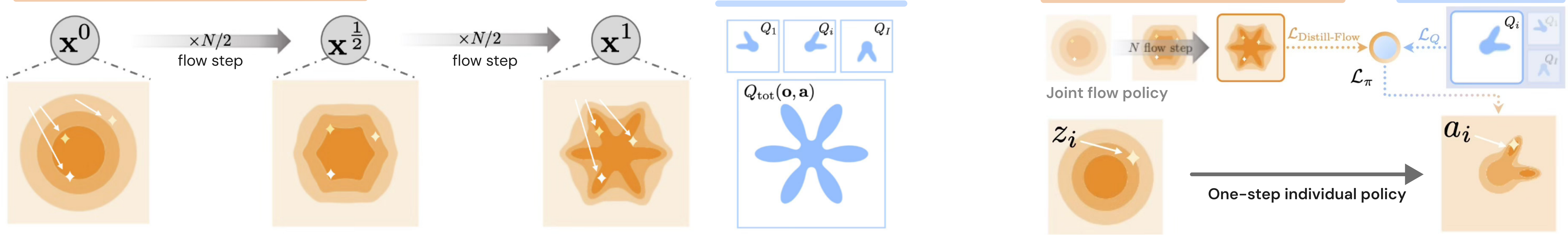
Questions are here: how to

1. maintain **decentralized manner** in deployment?
2. maximize **return/value function** in terms of **global level**?

MAC-Flow alleviates **performance-efficiency tradeoff** in offline MARL, achieving a **14.5x** speedup



MAC-Flow: First expressiveness, then efficiency



Joint flow policy extraction

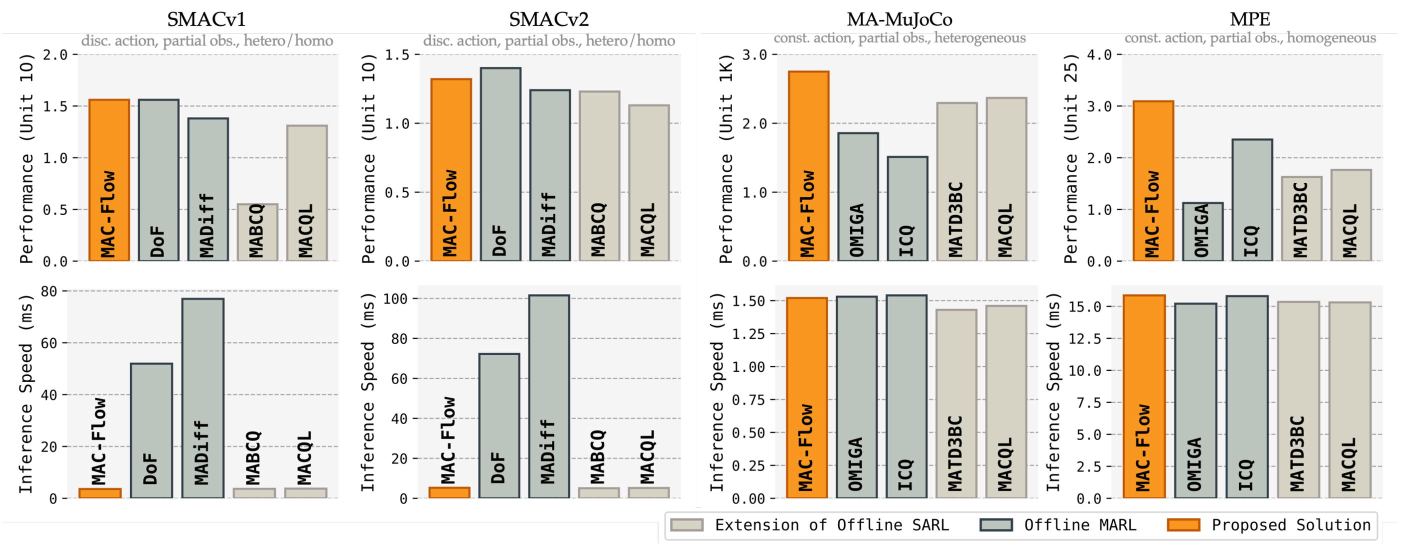
Flow policy factorization with Q maximization

Individual-global-max principle

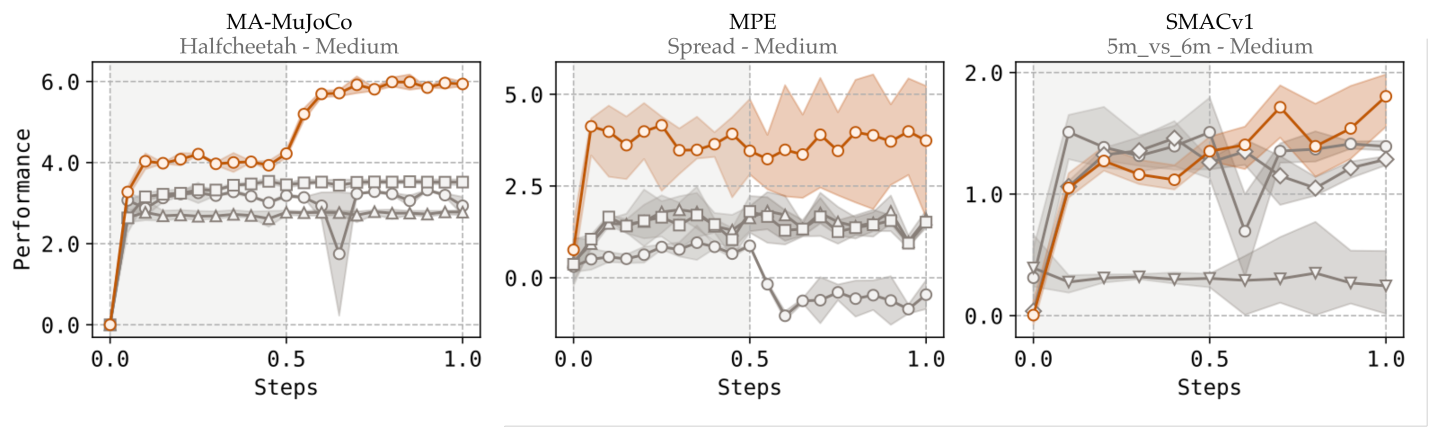
Property of MAC-Flow

Experimental results

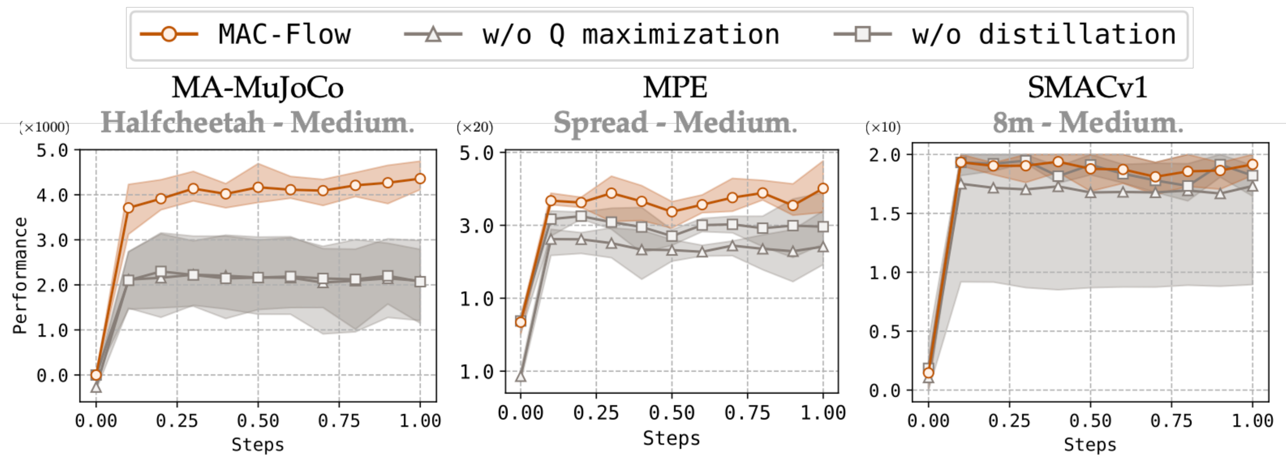
Offline MARL across continuous and discrete action space

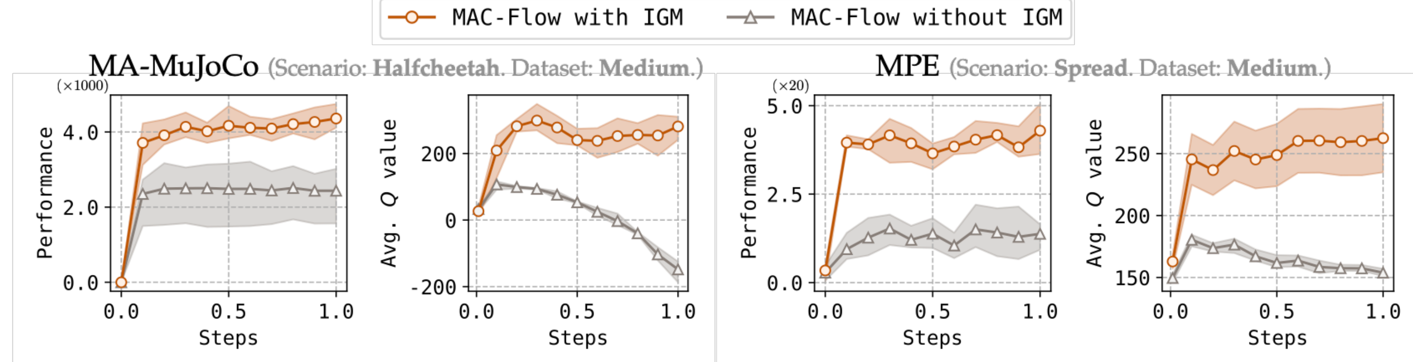


Offline-to-online MARL



Ablation studies





Distillation with RL objectives and IGM-based critics are the core drivers of MAC-Flow's performance and stability