Deep Meta Coordination Graphs for Multi-Agent Reinforcement Learning

Extended Abstract

Nikunj Gupta
University of Southern California
Los Angeles, CA, USA
nikunj@usc.edu

James Zachary Hare
DEVCOM Army Research Laboratory
Adelphi, MD, USA
james.z.hare.civ@army.mil

Jesse Milzman
DEVCOM Army Research Laboratory
Adelphi, MD, USA
jesse.m.milzman.civ@army.mil

Rajgopal Kannan
DEVCOM ARL Army Research Office
Los Angeles, CA, USA
rajgopal.kannan.civ@army.mil

Viktor Prasanna
University of Southern California
Los Angeles, CA, USA
prasanna@usc.edu

ABSTRACT

This paper presents *deep meta coordination graphs* (DMCG) for learning cooperative policies in multi-agent reinforcement learning (MARL). Coordination graph formulations encode local interactions and accordingly factorize the joint value function of all agents to improve efficiency in MARL. DMCG learns a more expressive representation of agent interactions and use them to integrate agent information through graph convolutional networks. The goal is to enable an evolving coordination graph to guide effective coordination in cooperative MARL tasks. The graphs are jointly optimized with agents' value functions to learn to implicitly reason about joint actions, facilitating the end-to-end learning of interaction representations and coordinated policies. We demonstrate that DMCG consistently achieves state-of-the-art coordination performance and sample efficiency on challenging cooperative tasks, outperforming several prior graph-based and non-graph-based MARL baselines. Through several ablations, we also isolate the impact of individual components in DMCG, showing that the observed improvements are due to the meaningful design choices in this approach. All codes can be found here: https://github.com/Nikunj-Gupta/dmcg-marl.

KEYWORDS

Multi-agent reinforcement learning; coordination; GNNs

ACM Reference Format:
Nikunj Gupta, James Zachary Hare, Jesse Milzman, Rajgopal Kannan, and Viktor Prasanna. 2026. Deep Meta Coordination Graphs for Multi-Agent Reinforcement Learning: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/IIKC6491

1 INTRODUCTION

Multi-agent reinforcement learning (MARL) has become essential for cooperative tasks such as warehouse robotics, drone swarms, and autonomous vehicles, where agents must coordinate under partial observability [1, 3, 7, 10, 12, 17, 27]. Value-decomposition methods [8, 9, 20–24] scale training via centralized training with decentralized execution, but their fully factorized value functions struggle with credit assignment and *relative overgeneralization* [2, 18, 25], where an agent's action may appear uninformative because its payoff depends on whether others coordinate. Coordination graphs (CGs) [2, 4, 5, 11, 14, 15, 19, 25] address this by decomposing the joint value function into individual utilities and pairwise payoff terms, enabling agents to reason about how subsets of teammates affect outcomes. Deep coordination graphs (DCG) [2] extend this idea to deep MARL with fixed topologies, while subsequent work has explored implicit coordination via attention (DICG) [15], richer edge modeling [11, 25], and dynamic agent grouping [4, 19]. However, effective coordination often requires reasoning about *multiple types* of dynamically evolving dependencies, such as direct interactions, implicit communication links, and influence-based relationships, that are not known *a priori* and must be discovered during learning.

We propose **deep meta coordination graphs (DMCG)**, which dynamically compose what we call *meta coordination graphs* (MCGs) to learn expressive representations of agent interactions. MCGs are integrated with graph convolutional networks (GCNs) [13, 26] for agent information aggregation and jointly optimized with factored Q-value functions in a fully differentiable, end-to-end pipeline. A full description of the approach, including extended experiments and ablation studies, is available in the full paper [6].

2 DEEP META COORDINATION GRAPHS

Figure 1 illustrates our methodology. We model cooperative tasks as a decentralized partially observable Markov decision process (Dec-POMDP) [16] with n agents. We maintain K base relation graphs represented as adjacency matrices $\{A_k\}_{k=1}^{K}$, each encoding a distinct latent interaction type among agents and initialized as complete graphs to maximize expressivity. Agent observations are stacked into a feature matrix $X = [o_1; \dots; o_n] \in \mathbb{R}^{n \times d}$. Through L attention-based composition layers, DMCG constructs C parallel channels, where each layer forms a soft mixture of the base graphs as $A^{(\ell,c)} = \sum_{k=1}^{K} \alpha_k^{(\ell,c)} A_k$, with attention weights parameterized by trainable matrices $W^{(\ell)} \in \mathbb{R}^{C \times K}$. Sequentially composing these layers yields a channel-specific meta coordination graph

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/IIKC6491

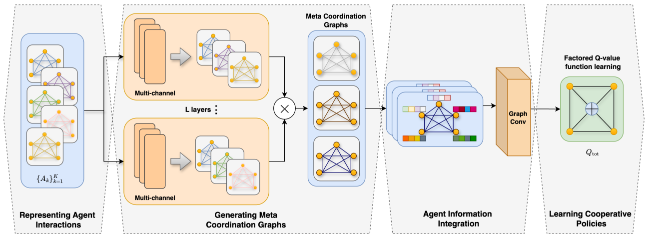

Figure 1: Overview of Deep meta coordination graphs (DMCG). From left to right: (1) agent interactions are represented through a set of base relation graphs, (2) multiple attention-based composition layers combine these into task-adaptive *meta coordination graphs* (MCGs), (3) MCGs guide graph convolutions for agent information integration, and (4) the resulting embeddings are used for factored Q-value learning to produce cooperative policies. Together, these steps enable DMCG MARL agents to achieve robust coordination and strong sample efficiency across challenging tasks.

(MCG) $A_M^{(c)} = \prod_{\ell=1}^{L} A^{(\ell,c)}$, enabling the capture of multi-hop relational dependencies without explicitly introducing long-range edges, while the C channels maintain diverse coordination hypotheses. Each MCG then guides a graph convolution [13] with self-loops, $H^{(c)} = \sigma(\tilde{D}_c^{-1}\tilde{A}_M^{(c)}XW)$ with $\tilde{A}_M^{(c)} = A_M^{(c)} + I$, and the resulting channel-wise outputs are concatenated and projected to obtain per-agent embeddings $Z \in \mathbb{R}^{n\times(C\cdot d_{\text{emb}})}$. These embeddings replace raw observations in a coordination-graph-based value factorization (following DCG [2]), decomposing Q_{tot} into individual utilities Q_i and pairwise payoffs Q_{ij} over a complete graph, with MCG parameters and value functions optimized jointly end-to-end. This design allows DMCG to capture dynamic interaction patterns while preserving the theoretical guarantees of CG-based factorization.

3 EXPERIMENTS AND RESULTS

We evaluate on four tasks from the MACO benchmark [25], designed to stress-test coordination under temporal extension, stochastic dynamics, and penalties for miscoordination: `Gather` (5 agents navigating to a shared goal), `Disperse` (12 agents distributing across hospitals), `Pursuit` (10 predators coordinating to capture prey), and `Hallway` (multi-group synchronization through corridors). We compare against VDN [23], QMIX [21], DCG [2], DICG [15], CASEC [11], NLCG [25], GACG [4], and VAST [19] across four seeds per task.

Main results (Figure 2). DMCG consistently achieves the best or near-best performance across all tasks. In Gather, it reaches ~98% win rate within ~180K episodes, far surpassing DCG (~80%) and outpacing DICG. In Disperse, DMCG learns markedly faster while reaching competitive final scores. On the more challenging Pursuit and Hallway, where miscoordination incurs severe penalties, DMCG achieves clear advantages in both convergence speed and final score over all baselines.

Ablations (Figure 3). Targeted ablations on Gather confirm each component's importance: (i) reducing composition layers (L=1) causes a sharp performance drop, confirming multi-layer refinement is critical; (ii) a single channel (C=1) converges to lower win

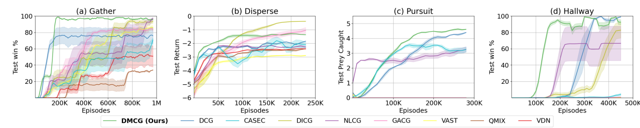

Figure 2: Overall performance. Comparison of DMCG with representative MARL baselines across four tasks: `Gather`, `Disperse`, `Pursuit`, and `Hallway`. DMCG consistently achieves state-of-the-art or near-optimal performance, converging faster and attaining higher mean episode returns than static graph (DCG), attention-based graph (DICG), edge-selection (CASEC), non-linear mixers (NLCG), subgrouping (GACG, VAST), and value-decomposition methods (QMIX, VDN).

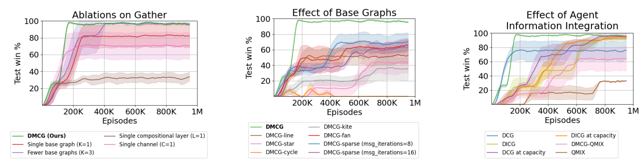

Figure 3: Ablation analysis on *Gather*. *Left:* Targeted ablations of DMCG varying composition depth (L), channels (C), and base relation bank size (K). *Middle:* Effect of base graph initialization, comparing fully connected and sparse graphs; increasing message passing (8→16) does not close the gap. *Right:* Capacity scaling in DCG and DICG versus MCG-based integration; naive scaling underperforms DMCG, while MCG-based integration yields consistent gains.

rates, showing that parallel channels help explore alternative interaction patterns; (iii) smaller base relation banks (K=1 or K=$\lceil n/2\rceil$) limit diversity and sample efficiency. We also find that fully connected base graph initialization outperforms sparse alternatives (line, star, cycle, kite), though diverse sparse mixtures perform reasonably. Importantly, simply scaling DCG (wider payoff networks) or DICG (deeper GCNs) to comparable parameter counts does not match DMCG, confirming that gains stem from its compositional MCG design rather than model size.

4 CONCLUSION AND FUTURE WORK

DMCG advances cooperative MARL by dynamically composing task-adaptive meta coordination graphs that capture rich, evolving agent interactions. Its compositional design, including multi-layer refinement, parallel channels, and diverse base relations, enables expressive agent information integration that outperforms existing CG-based and value-decomposition methods. Future work includes scaling to larger agent populations, mixed cooperative-competitive settings, and integrating domain priors for further efficiency.

ACKNOWLEDGMENTS

This work is supported by DEVCOM ARL Army Research Office; Grant: W911NF2420194; and U.S. National Science Foundation (NSF); Grant: OAC-2411446. Distribution Statement A: Approved for public release. Distribution is unlimited.

REFERENCES

[1] Afnan M Alharbi, Ghaida Alshehri, and Salma Elhag. 2024. Reinforcement learning of emerging swarm technologies: A literature review. In *Proceedings of the Future Technologies Conference*. Springer, 478–494.

[2] Wendelin Böhmer, Vitaly Kurin, and Shimon Whiteson. 2020. Deep coordination graphs. In *International Conference on Machine Learning*. PMLR, 980–991.

[3] Joris Dinneweth, Abderrahmane Boubezoul, René Mandiau, and Stéphane Espié. 2022. Multi-agent reinforcement learning for autonomous vehicles: A survey. *Autonomous Intelligent Systems* 2, 1 (2022), 27.

[4] Wei Duan, Jie Lu, and Junyu Xuan. 2024. Group-Aware Coordination Graph for Multi-Agent Reinforcement Learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, Kate Larson (Ed.). International Joint Conferences on Artificial Intelligence Organization, 3926–3934. https://doi.org/10.24963/ijcai.2024/434 Main Track.

[5] Carlos Guestrin, Michail G Lagoudakis, and Ronald Parr. 2002. Coordinated Reinforcement Learning. In *Proceedings of the Nineteenth International Conference on Machine Learning*. 227–234.

[6] Nikunj Gupta, James Zachary Hare, Jesse Milzman, Rajgopal Kannan, and Viktor Prasanna. 2026. Deep Meta Coordination Graphs for Multi-agent Reinforcement Learning. arXiv:2502.04028 [cs.LG] https://arxiv.org/abs/2502.04028

[7] Nikunj Gupta, G Srinivasaraghavan, Swarup Mohalik, Nishant Kumar, and Matthew E Taylor. 2025. Hammer: Multi-level coordination of reinforcement learning agents via learned messaging. *Neural Computing and Applications* 37, 19 (2025), 13221–13236.

[8] Nikunj Gupta, Ludwika Twardecka, James Zachary Hare, Jesse Milzman, Rajgopal Kannan, and Viktor Prasanna. 2025. TIGER-MARL: Enhancing Multi-Agent Reinforcement Learning with Temporal Information through Graph-based Embeddings and Representations. arXiv:2511.08832 [cs.LG] https://arxiv.org/abs/2511.08832

[9] Tarun Gupta, Anuj Mahajan, Bei Peng, Wendelin Böhmer, and Shimon Whiteson. 2021. Uneven: Universal value exploration for multi-agent reinforcement learning. In *International Conference on Machine Learning*. PMLR, 3930–3941.

[10] Pablo Hernandez-Leal, Bilal Kartal, and Matthew E Taylor. 2019. A survey and critique of multiagent deep reinforcement learning. *Autonomous Agents and Multi-Agent Systems* 33, 6 (2019), 750–797.

[11] Yipeng Kang, Tonghan Wang, Qianlan Yang, Xiaoran Wu, and Chongjie Zhang. 2022. Non-linear coordination graphs. *Advances in Neural Information Processing Systems* 35 (2022), 25655–25666.

[12] Russell Keith and Hung Manh La. 2024. Review of autonomous mobile robots for the warehouse environment. *arXiv preprint arXiv:2406.08333* (2024).

[13] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SJU4ayYgl

[14] Jelle R Kok and Nikos Vlassis. 2006. Collaborative multiagent reinforcement learning by payoff propagation. *Journal of machine learning research* 7 (2006).

[15] Sheng Li, Jayesh K Gupta, Peter Morales, Ross Allen, and Mykel J Kochenderfer. 2021. Deep Implicit Coordination Graphs for Multi-agent Reinforcement Learning. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 764–772.

[16] Frans A Oliehoek, Christopher Amato, et al. 2016. *A concise introduction to decentralized POMDPs*. Vol. 1. Springer.

[17] Afshin Oroojlooy and Davood Hajinezhad. 2023. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence* 53, 11 (2023), 13677–13722.

[18] Liviu Panait, Sean Luke, and R Paul Wiegand. 2006. Biasing coevolutionary search for optimal multiagent behaviors. *IEEE Transactions on Evolutionary Computation* 10, 6 (2006), 629–645.

[19] Thomy Phan, Fabian Ritz, Lenz Belzner, Philipp Altmann, Thomas Gabor, and Claudia Linnhoff-Popien. 2021. Vast: Value function factorization with variable agent sub-teams. *Advances in Neural Information Processing Systems* 34 (2021), 24018–24032.

[20] Tabish Rashid, Gregory Farquhar, Bei Peng, and Shimon Whiteson. 2020. Weighted qmix: Expanding monotonic value function factorisation for deep multi-agent reinforcement learning. *Advances in neural information processing systems* 33 (2020), 10199–10210.

[21] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. In *International Conference on Machine Learning*. PMLR, 4295–4304.

[22] Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. 2019. Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning. In *International conference on machine learning*. PMLR, 5887–5896.

[23] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2018. Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 2085–2087.

[24] Jianhao Wang, Zhizhou Ren, Terry Liu, Yang Yu, and Chongjie Zhang. 2021. {QPLEX}: Duplex Dueling Multi-Agent Q-Learning. In *International Conference on Learning Representations*. https://openreview.net/forum?id=Rcmk0xxIQV

[25] Tonghan Wang, Liang Zeng, Weijun Dong, Qianlan Yang, Yang Yu, and Chongjie Zhang. 2022. Context-Aware Sparse Deep Coordination Graphs. In *International Conference on Learning Representations*. https://openreview.net/forum?id=wQfgfb8VKTn

[26] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* 32, 1 (2020), 4–24.

[27] Ruiqi Zhang, Jing Hou, Florian Walter, Shangding Gu, Jiayi Guan, Florian Röhrbein, Yali Du, Panpan Cai, Guang Chen, and Alois Knoll. 2024. Multi-agent reinforcement learning for autonomous driving: A survey. *arXiv preprint arXiv:2408.09675* (2024).