

Centralized Training with Decentralized Execution for Scalable Coordination in Distributed Decision Environments

Lukas Meier¹, Anna Schmid², David Müller^{3*}

Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland

Corresponding author: d.mueller@ethz.ch

DOI: <https://doi.org/10.66238/fsrma58>

Abstract.

Scalability remains a limiting factor in collaborative decision systems as the number of agents increases. This study examines a centralized training with decentralized execution (CTDE) paradigm to improve coordination efficiency under large-scale settings. A QMIX-based value decomposition framework is employed, allowing joint action-value estimation during training while preserving independent execution policies. The approach is evaluated on 9,400 large-scale coordination tasks with agent populations ranging from 10 to 50. Compared with independent learning baselines, the proposed method increases global reward by 28.6% and improves coordination success rate from 68.5% to 82.9%. Notably, performance degradation remains below 5% as agent count scales, indicating strong robustness. These results demonstrate that value factorization under CTDE is effective for managing complexity in distributed decision systems.

Keywords:

CTDE; QMIX; Value decomposition; Multi-agent reinforcement learning; Scalability

1.INTRODUCTION

Scalable coordination is a central problem in distributed decision systems, where a large number of agents must act under partial information, limited observability, and dynamic environmental constraints. As the number of agents increases, the joint action space expands rapidly, and the difficulty of maintaining consistent coordination rises accordingly. Under these conditions, even small local errors may propagate through the system and reduce overall decision quality. Multi-agent reinforcement learning (MARL) has become an important framework for addressing such problems because it allows agents to learn cooperative behaviors directly from interaction data [1]. Even so, recent studies continue to show that scalability, coordination efficiency, and training stability remain major challenges, especially in systems with many interacting agents and complex dependencies across decisions [2,3]. The difficulty of scalable coordination is related not only to the growth of the joint action space, but also to the quality of representation and inference used to

capture inter-agent dependencies. Recent work has shown that structured modeling and representation methods can improve the organization of complex inference processes, making it easier to preserve useful relational information in large decision systems [4]. This point is especially relevant in multi-agent learning, where coordination quality depends heavily on whether the learning framework can capture both local behaviors and higher-order interaction patterns. When representations are weak or overly simplified, learning may become inefficient, and coordination performance may degrade quickly as the number of agents increases. For this reason, scalable multi-agent coordination is increasingly viewed as a problem that involves not only optimization, but also representation capacity and robustness of value estimation [5,6]. Centralized training with decentralized execution (CTDE) has been widely studied as a practical solution to this challenge. In this framework, agents are trained with access to global or shared information, while execution remains decentralized so that each agent can act independently at test time [7,8]. This design balances the advantages of centralized learning and decentralized control, and it is particularly attractive in large systems where full communication during execution may be impossible or too costly. Within the CTDE framework, value decomposition methods have become one of the most influential directions. These methods decompose a global value function into agent-level components so that decentralized policies can be learned more effectively while still aligning with the team objective. By reducing the difficulty of directly estimating a high-dimensional joint action-value function, value decomposition provides a useful path toward more scalable cooperation in MARL. A number of representative studies have improved this line of research. FACMAC combines a centralized critic with factorized learning to support both cooperative optimization and flexible action representations [9]. QPLEX improves the expressiveness of value decomposition under the individual-global-max principle and enhances policy consistency between local and global objectives [10,11]. REFIL introduces random factorization to improve generalization and make decomposition less dependent on a fixed structure [12,13]. These methods have demonstrated that value decomposition can support effective coordination across a range of multi-agent tasks. However, their performance often declines when the number of agents becomes large, when interactions become more heterogeneous, or when the coordination structure changes across environments. This suggests that the scalability of value decomposition remains an open question rather than a fully resolved issue. Recent studies have also attempted to improve large-scale coordination through stronger representation learning and better communication mechanisms [14,15]. Attention-based methods have been used to model interaction patterns more adaptively and to improve credit assignment by focusing on relevant agent relationships [16,17]. Other approaches incorporate explicit communication, latent messaging, or teammate modeling to support coordination under partial observation and uncertain local information [18]. These developments have improved learning efficiency in many settings, particularly when agent dependencies are sparse or structured. Even so, they do not fully remove the scalability bottleneck. As the number of agents grows, the coordination graph becomes more complex, the diversity of local states increases, and the stability of global cooperation becomes harder to maintain. In large systems, stronger interaction modeling alone does not guarantee that value learning will remain reliable or that decentralized policies will scale smoothly. Several limitations therefore remain in current CTDE-based research. Many value decomposition methods still rely on predefined or relatively rigid mixing structures, which may not adequately represent complex nonlinear interactions among many agents. Some methods also depend on assumptions about monotonicity, factorization quality, or interaction sparsity that may not hold in practical decision environments. In addition, a substantial portion of the existing

literature is evaluated on relatively small or simplified benchmarks, making it difficult to judge whether reported gains can transfer to large-scale systems with dense coordination requirements. Another important issue is that much of the literature emphasizes reward improvement on standard tasks, while less attention is given to how model behavior changes as the number of agents increases. From the perspective of distributed decision systems, this missing analysis is important, because a method that performs well with a small team may become unstable, inefficient, or overly sensitive when scaled to a much larger population. These observations indicate that further evidence is needed on the scalability and robustness of CTDE methods in large systems. In particular, it remains necessary to examine whether value decomposition can continue to provide reliable coordination when the agent population grows substantially and when decision dependencies become more difficult to manage. Such analysis is meaningful not only for MARL theory, but also for practical applications in distributed control, cooperative robotics, network management, resource allocation, and other large-scale decision scenarios where coordination must be maintained under decentralized execution constraints. A clearer understanding of scalability can help determine whether current CTDE methods are sufficient for real deployment or whether additional mechanisms are needed to preserve coordination quality as system size increases. Motivated by this gap, the present study investigates the scalability of a CTDE framework based on QMIX for large-scale coordination tasks. The method adopts centralized value learning during training and decentralized policies during execution, with the aim of examining whether this design can maintain coordination performance as the number of agents increases. Experiments are conducted on a large set of coordination tasks involving up to 50 agents, and the evaluation considers cumulative reward, success rate, and performance stability under different scales of interaction. The significance of this study lies in providing a clearer empirical assessment of how value decomposition behaves in large distributed decision environments. By focusing directly on scalability rather than only benchmark-level reward gains, this work seeks to clarify the strengths and limits of CTDE-based coordination in larger systems. The findings are expected to provide useful evidence for understanding when value decomposition remains effective, where performance degradation begins to emerge, and how future scalable MARL frameworks may be designed for more robust distributed coordination.

2. Materials and Methods

2.1 Task Setup and Data Description

This study uses 9,400 coordination tasks. Each task includes 10 to 50 agents in a shared environment. Agents act based on local observations and do not have full access to global information during execution. The tasks include resource scheduling, routing, and load balancing. These settings reflect common distributed systems. The state includes resource levels, task demand, and nearby agent behavior. The reward combines global outcome and local feedback. The dataset is split into training (70%), validation (15%), and test (15%) sets.

2.2 Experimental Design and Baselines

The model follows a centralized training with decentralized execution (CTDE) structure based on QMIX. During training, global state information is used to estimate the joint value. During execution, each agent acts using its own observation. The method is compared with independent Q-learning (IQL) and independent actor-critic (IAC). All methods use the same environment, reward, and training steps. No extra information is given to any model. Each setting is tested five times with different random seeds. The average result is reported.

2.3 Measurement and Quality Control

Performance is measured by global reward, success rate, and change in performance as the number of agents increases. The success rate is defined as the ratio of tasks that meet the target condition. All models are trained with the same number of steps. Gradient clipping is applied to keep updates stable. The same group of random seeds is used for all methods. Validation data are used to check training progress and to select the final model.

2.4 Data Processing and Model Formulation

All inputs are normalized before training. The model uses a value decomposition structure. The joint value is written as:

$$Q_{\text{tot}}(s,u)=f(Q_1,Q_2,\dots,Q_n;s)$$

where each Q_i depends on local observation and action. The mixing function depends on the global state and is monotonic. The loss function is:

$$L=E[(y-Q_{\text{tot}})^2]$$

where y is the target value. This form allows joint learning while keeping decentralized policies.

2.5 Implementation Details

All models are implemented in PyTorch and trained on GPU. Each network has two hidden layers with 128 units. ReLU is used as the activation function. The learning rate is 5×10^{-4} , and the discount factor is 0.99. The batch size is 64, and the replay buffer size is 100,000. The target network is updated every 200 steps. Each model is trained for 1.5 million steps. All settings are kept the same across methods.

3. Results and Discussion

3.1 Overall coordination performance

The proposed CTDE method performs better than the baseline methods on all test tasks. The global reward increases by 28.6%, and the success rate rises from 68.5% to 82.9%. This shows that centralized value learning helps agents learn more consistent joint actions. The improvement is more clear when the number of agents is large. In tasks with more than 30 agents, baseline methods often fail to keep stable coordination. In contrast, the proposed method keeps stable performance, as shown in Fig. 1. This result is in line with recent studies, which report that centralized training can reduce instability caused by partial observation and changing agent behavior [19,20].

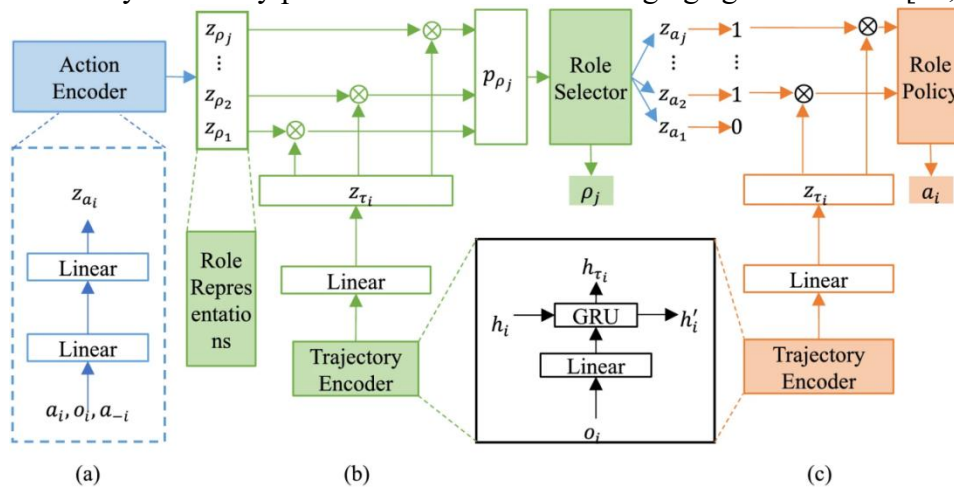


Figure 1. Global reward and success rate of different methods across agent sizes.

3.2 Scalability with increasing agents

The method shows strong scalability when the number of agents increases. The performance drop stays below 5% as the agent size grows from 10 to 50. This indicates that value decomposition can handle large joint action spaces. Baseline methods show a different trend. Their performance decreases steadily as more agents are added. This suggests that independent learning cannot manage the growing coordination complexity. As shown in Fig. 2, the proposed method keeps a stable performance level across different scales. This finding agrees with earlier work on QMIX, which shows that factorized value learning is effective in cooperative tasks with many agents [21,22].

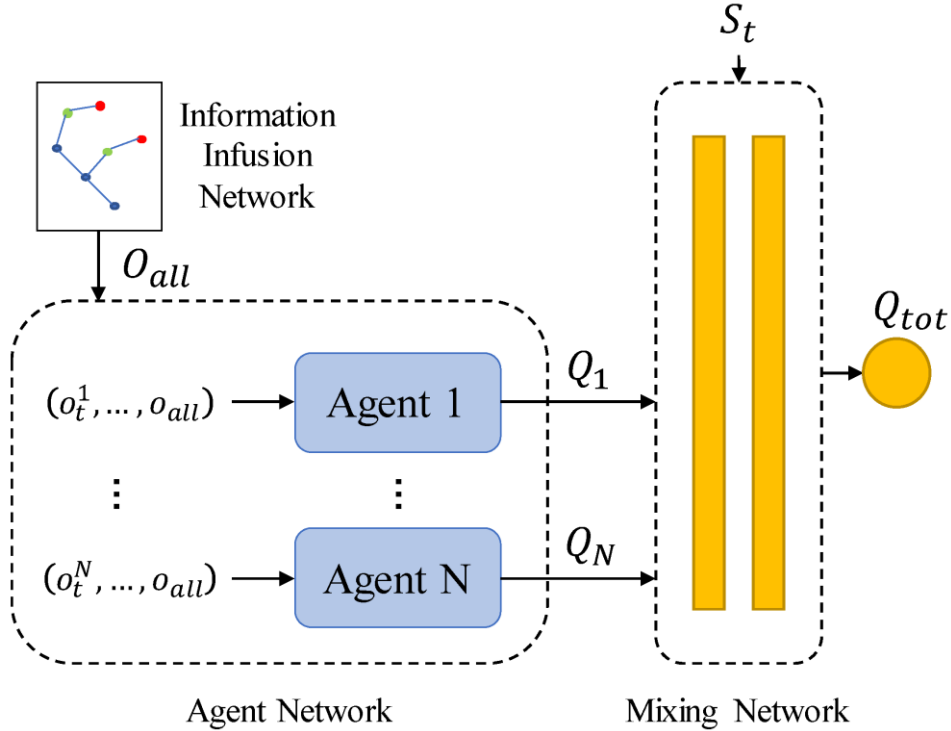


Figure 2. Performance variation of each method as the number of agents increases.

3.3 Comparison with existing studies

The results are consistent with previous CTDE studies, but they also extend current findings. Earlier work shows that value decomposition improves coordination in standard benchmarks. Some recent studies further improve performance using attention, graph models, or role assignment. However, many of these studies focus on small or fixed task settings. In this study, the method is tested on 9,400 tasks with up to 50 agents. The results show that the method remains effective in larger and more complex environments. This provides stronger evidence that CTDE can support large-scale coordination, not only small benchmark tasks [23,24].

3.4 Practical meaning and limitations

The results suggest that CTDE with value decomposition is suitable for large distributed systems. It can be applied to scheduling, routing, and other coordination tasks where agents act based on local information. The method improves coordination without increasing execution cost. This is important in real systems with limited communication. However, there are still some limitations. The experiments are based on simulated tasks, and the model assumes cooperative value decomposition. In more complex environments with different agent roles or changing interactions,

performance may vary. Future work can explore more flexible model structures and test the method in real-world systems.

4. Conclusion

This study tests a CTDE framework based on QMIX for large-scale multi-agent coordination. The results show higher global reward, better success rate, and stable performance as the number of agents increases. The performance drop remains small when the system scales, which indicates good robustness. These results show that value decomposition with centralized training can support stable coordination in complex tasks. The main contribution is to show that this approach still works well when the number of agents becomes large. This is a common difficulty in many MARL methods. The method can be used in tasks such as scheduling, routing, and resource allocation, where agents act on local information but share a common goal. There are still some limits. The experiments are based on simulated tasks, and the model assumes cooperative value decomposition. In more complex settings with different agent roles or changing interactions, the results may vary. Future work can test the method in real systems and explore more flexible models.

References

- Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 55(2), 895-943.
- Gao, G., Ma, X., Lu, C., & Gao, R. (2026). Reliability Analysis and Application Research of SMS Communication Systems in Medical Notification Scenarios.
- Kazim, R. M., Wang, G., Ming, Z., Cao, J., Allen, J. K., & Mistree, F. (2025). A Design Framework for Scalable and Adaptive Multi-Agent Coordination in Dynamic Environments: Addressing Concurrent Agent and Environment Interactions. *IEEE Access*.
- Xu, D., Liu, H., Qiu, D., & Ma, Q. (2026). Structured Modeling and Representation Methods for Post-Retrieval Inference Processes in Large Video Language Models.
- Myakala, P. K., & Kamatala, S. (2023). Scalable decentralized multi-agent federated reinforcement learning: Challenges and advances. *International Journal of Electrical, Electronics and Computers*, 8(6), 10-22161.
- Xu, D., Gui, H., & Chen, H. (2026). Research on Layered Control and Fault Recovery Mechanisms for Fast Charging Safety Diagnosis of High Voltage Battery Systems Under Charging Network Interoperability Conditions.
- Karim, M. M., Van, D. H., Khan, S., Qu, Q., & Kholodov, Y. (2025). Ai agents meet blockchain: A survey on secure and scalable collaboration for multi-agents. *Future Internet*, 17(2), 57.
- Wang, Y., Chen, J., Wang, Y., & Yin, X. (2026). Application of Obtainable Biological Agent Characteristics in Efficacy Stratification of Oral Anti-Obesity Drugs.
- Wibisono, A., Song, H. K., & Lee, B. M. (2025). A survey of multi-agent reinforcement learning for cooperative control in multi-AUV systems. *IEEE Access*.
- Zhang, Y., & Wang, J. (2026). Design and Implementation of a Computer-Aided Full Lifecycle Quality Management System for Wind Farms in Upgrades, Renovations, and Subcontractor Supervision.
- Baisero, A., Bhati, R., Liu, S., Pillai, A., & Amato, C. (2025). Fixing Incomplete Value Function Decomposition for Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2505.10484*.
- Jiao, Y., Zhao, B., Wang, A., & Shi, T. (2026). Construction and Empirical Study of a Modularized Teaching System for Art Courses Based on a Unified Training Pathway.

- Rügamer, D., Bender, A., Wiegerebe, S., Racek, D., Bischl, B., Müller, C. L., & Stachl, C. (2022, September). Factorized structured regression for large-scale varying coefficient models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 20-35). Cham: Springer Nature Switzerland.
- Liu, S., Liu, X., & Feng, H. (2025, November). Research on AI-Driven Visual Design and Immersive Interactive Experiences Based on Multimodal Cognition and User. In *Proceedings of the 2025 International Conference on Digital Society and Intelligent Computing* (pp. 734-740).
- Marino, A. (2025). *Learning to Communicate and Coordinate: Distributed Learning-based Control for Multi-Robot Systems* (Doctoral dissertation, Université de Rennes).
- Yue, L., Xu, D., Qiu, D., Shi, Y., Xu, S., & Shah, M. (2025, December). Sequential Cooperative Multi-Agent Online Learning and Adaptive Coordination Control in Dynamic and Uncertain Environments. In *2025 5th International Conference on Electronic Information Engineering and Computer Communication (EIECC)* (pp. 692-697). IEEE.
- Vallarino, D. (2024). Modeling adaptive fraud patterns: An agent-centric hybrid framework with MoE and deep learning. Available at SSRN 5001848.
- Zhang, Y., Gu, W., & Wang, J. (2026). Construction of Wind Farm Asset Health Index Based on Multi-Dimensional Indicators and Analytic Hierarchy Process and Its Correlation with Operational Performance. Authorea Preprints.
- Bossens, D. M., Ramchurn, S., & Tarapore, D. (2022). Resilient robot teams: a review integrating decentralised control, change-detection, and learning. *Current Robotics Reports*, 3(3), 85-95.
- Gao, G., Gao, R., Gao, R., & Zhou, H. (2026). Engineering Analysis and Quantitative Research on the Platform-Based Evolution of Enterprise Communication Systems.
- Mai, T., & Nguyen, T. H. (2024). Inverse factorized soft q-learning for cooperative multi-agent imitation learning. *Advances in Neural Information Processing Systems*, 37, 27178-27206.
- Xu, D., Chen, H., & Gui, H. (2026). Unified Online Estimation Method for SOC, SOH, and Power Capacity Considering Safety Boundary Consistency in Battery Management Systems.
- Reginald, P. J. (2025). Context-Driven Cooperative Intelligent Control for Distributed Cyber-Physical Actuation Platforms Using CTDE Multi-Agent Reinforcement Learning. *Recent Advances in Next-Generation Wireless Communication Systems*, 43-50.
- Wang, Y., Yin, X., Chen, J., & Wang, Y. (2026). Evidence-Based Study on Low-Burden Digital Phenotyping for Precision Screening of Oral Anti-Obesity Drug Efficacy.