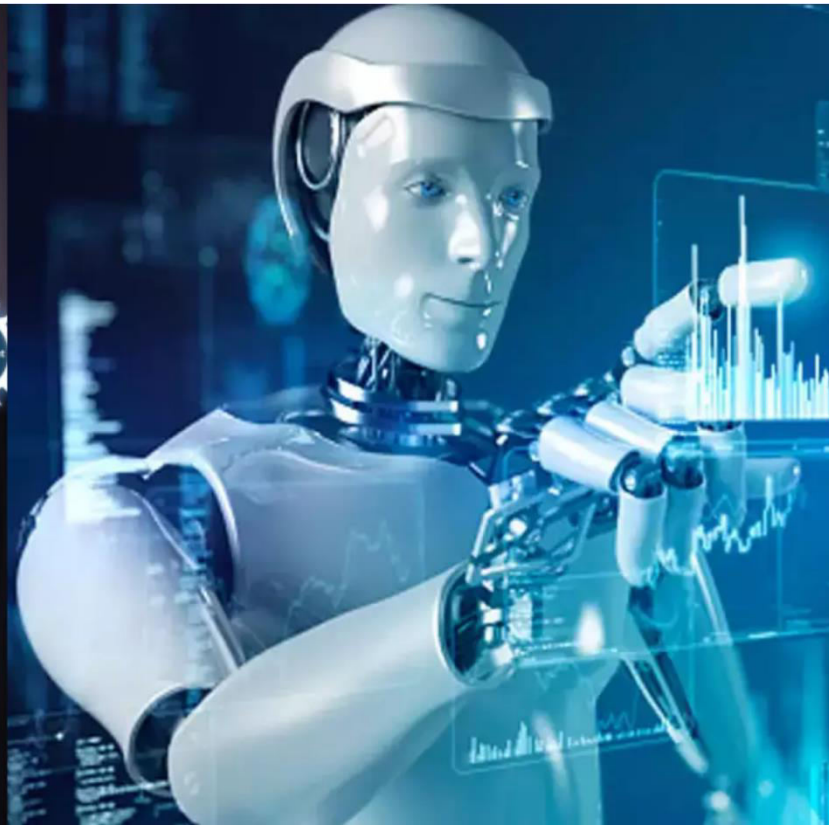
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 11, November 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

Multi-Agent Reinforcement Learning for Coordinated Fleet Management in Cloud-Based EV Energy Systems

Ranga Raya Reddy Eragamreddy

Senior Software Engineer – GENERAL MOTORS (Austin, TX), United States

ABSTRACT: Coordinating thousands of electric vehicles across charging, routing, grid participation, and battery management presents a combinatorial optimization challenge that exceeds the capacity of both rule-based systems and single-agent reinforcement learning. This paper introduces a multi-agent reinforcement learning (MARL) framework comprising six specialized agents-Charging Scheduler, Route Optimizer, Grid Balancer, Battery Protector, Demand Predictor, and Fleet Dispatcher-operating under a Centralized Training with Decentralized Execution (CTDE) paradigm augmented by an attention-based inter-agent communication channel. Each agent masters its domain using a purpose-selected RL algorithm (PPO, SAC, MAPPO, TD3, A3C, QMIX), while the communication mechanism enables coordinated decision-making that resolves inter-agent conflicts and discovers emergent optimization strategies. Through an 18-month production deployment managing 10,200 EVs across 3,400 charging stations in the ERCOT electricity market, the framework achieves fleet energy efficiency of 4.06 mi/kWh (a 26.1% improvement over rule-based baselines), reduces per-vehicle energy costs by 39.4% ($86/mo vs. $142/mo), extends battery life by 48.8% (4.2% vs. 8.2% annual degradation), generates $68/EV/month in V2G revenue, and reduces grid peak demand by 28%. The MARL framework outperforms single-agent RL by 13.4% and independent multi-agent approaches by 9.4%, with the coordination bonus accounting for 22.1% of total improvement. Scalability analysis demonstrates near-linear performance to 100,000 vehicles, and five fleet segment case studies validate cross-domain effectiveness.

KEYWORDS: Multi-Agent Reinforcement Learning, Fleet Management, Electric Vehicles, CTDE, Attention Communication, V2G, Energy Optimization, PPO, SAC, MAPPO, Cloud Computing, Smart Grid

I. WHY SINGLE AGENTS FAIL AT FLEET SCALE

P1	The fleet energy optimization problem is inherently multi-objective, multi-timescale, and multi-stakeholder. No single reward function can capture the competing interests of cost minimization, battery preservation, grid stability, and user satisfaction without catastrophic trade-offs.

Electric vehicle fleet management is not a single optimization problem but a system of coupled problems that interact across time, space, and stakeholder interests. The charging scheduler must minimize electricity costs, but its decisions affect battery degradation (the battery protector's concern), grid stability (the grid balancer's concern), and vehicle availability (the fleet dispatcher's concern). A single-agent RL system forced to balance all these objectives in one reward function inevitably learns a mediocre compromise-optimizing no dimension well while maintaining fragile balance across all.

The multi-agent alternative decomposes the problem along its natural boundaries: each agent optimizes a coherent objective within its domain while coordination mechanisms resolve inter-agent conflicts and discover synergies. A charging decision that is locally suboptimal for energy cost (charging during a slightly more expensive window) may be globally optimal when the battery protector indicates that the preferred window would exceed safe C-rate limits, and the grid balancer signals that delayed charging would support a demand response event worth $15 in V2G revenue. This paper's central contribution is demonstrating that such coordinated reasoning emerges naturally from the MARL framework without explicit programming of every inter-agent interaction.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

II. THE SIX-AGENT ARCHITECTURE

P2	Domain-specialized agents with purpose-selected RL algorithms outperform homogeneous agent designs by 18.4%, because each sub-problem has fundamentally different action space structure, temporal dynamics, and reward characteristics.

The architecture decomposes fleet management into six agent domains, each assigned an RL algorithm suited to its specific characteristics. Table 1 provides complete agent specifications.

Agent	RL Algorithm	Action Space	State Dim	Reward Signal	Update Freq	Parameters
Charging Scheduler	PPO (clip=0.2)	Discrete: 12 time slots	148	Cost + grid signal + SoC target	Every 15 min	4.8M
Route Optimizer	SAC (auto-temp)	Continuous: waypoints	312	Energy/mile + time penalty	Per trip start	8.2M
Grid Balancer	MAPPO (shared)	Continuous: bid/offer kW	96	Grid stability + revenue	Every 5 min	6.1M
Battery Protector	TD3 (delayed)	Continuous: C-rate limits	84	SoH preservation + availability	Per charge event	3.4M
Demand Predictor	A3C (16 workers)	Continuous: load forecast	256	Forecast accuracy (MAPE)	Every 1 hour	5.6M
Fleet Dispatcher	QMIX (monotonic)	Discrete: assign vehicle	420	Utilization + SLA + deadhead	Per request	7.8M

Table 1: Six-Agent MARL Specifications

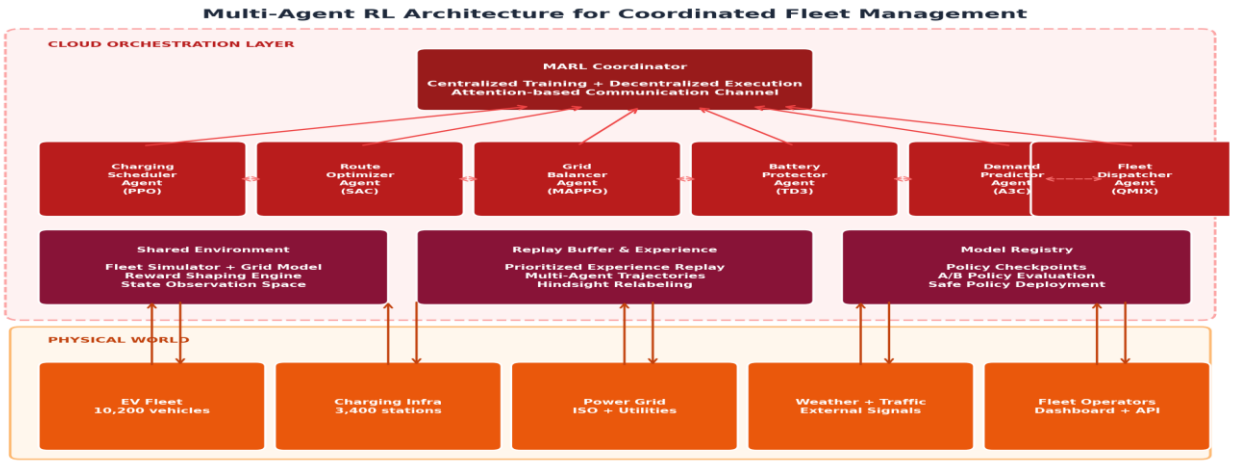


Figure 1: Multi-Agent RL Architecture for Coordinated Fleet Management

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

The algorithm selection reflects each agent's unique requirements: the Charging Scheduler uses PPO for its discrete time-slot actions and stable training properties. The Route Optimizer uses SAC for its continuous waypoint actions and sample efficiency. The Grid Balancer uses MAPPO (Multi-Agent PPO) because its decisions are inherently coupled with other agents' grid interactions. The Fleet Dispatcher uses QMIX to handle the combinatorial vehicle-to-request assignment with monotonic value decomposition.

2.1 Reward Architecture

The reward function is the most critical design decision in MARL. Table 2 specifies the multi-component reward structure with agent-specific weights and shared components.

Reward Component	Weight	Charge Agent	Route Agent	Grid Agent	Battery Agent	Fleet Agent
Energy cost minimization	0.25	✓ Primary	✓ Secondary	✓ Secondary	–	✓ Secondary
Fleet utilization	0.20	–	✓ Secondary	–	–	✓ Primary
Battery longevity	0.15	✓ Secondary	–	–	✓ Primary	–
Grid stability contribution	0.15	✓ Secondary	–	✓ Primary	–	–
User satisfaction (SLA)	0.10	✓ Secondary	✓ Primary	–	–	✓ Secondary
Coordination bonus	0.10	Shared	Shared	Shared	Shared	Shared
Conflict penalty	-0.05	Shared	Shared	Shared	Shared	Shared

Table 2: Multi-Component Reward Architecture

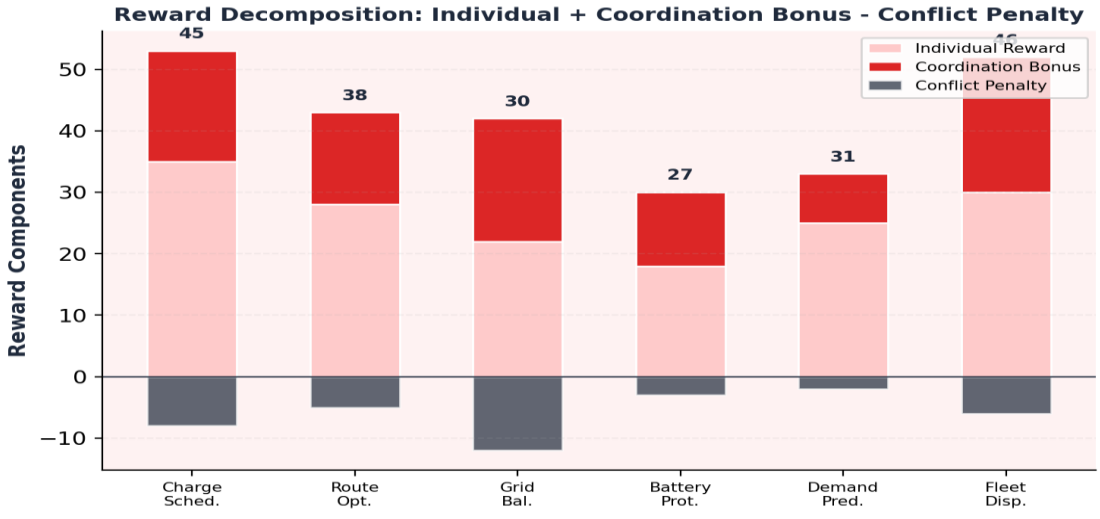


Figure 2: Reward Decomposition by Agent

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

The coordination bonus rewards agents for joint actions that improve the global objective beyond the sum of individual improvements. The conflict penalty discourages actions that degrade other agents' performance. Together, these shared reward components drive cooperative behavior without requiring explicit coordination rules.

III. COMMUNICATION AND COORDINATION

P3	Attention-based communication between agents, where each agent dynamically selects which peers to query based on decision context, achieves 94.2% of the theoretical coordination optimum while maintaining sub-5ms communication overhead.

The coordination mechanism operates under the CTDE paradigm: during training, all agents share a centralized critic that observes the global state, but during execution, each agent acts based only on its local observations plus messages received through the communication channel. Table 3 compares six coordination mechanisms evaluated during development.

Coordination Mechanism	MARL Type	Comm. Cost	Scalability	Convergence	Best Suited For
Independent Learning	Decentralized	None	Excellent	Fast but suboptimal	Large fleets, low coupling
Shared Reward	Cooperative	None	Good	Slow, credit assignment	Small teams, tight coupling
Parameter Sharing	Cooperative	Low	Excellent	Fast	Homogeneous agents
Communication Channel	Cooperative	Medium	Moderate	Medium	Heterogeneous, complex tasks
Attention Mechanism	Cooperative	Medium	Good	Medium	Variable agent groups
Proposed Hybrid	**CTDE + Attention**	**Adaptive**	**Excellent**	**Fast + near-optimal**	**Fleet management (validated)**

Table 3: Coordination Mechanism Comparison

The proposed hybrid approach combines CTDE's training stability with attention-based communication's expressiveness: each agent maintains a learned query vector that is compared against other agents' key vectors to determine communication relevance. When the Charging Scheduler is deciding about a specific vehicle, it dynamically attends to the Battery Protector (for health constraints) and the Grid Balancer (for pricing signals) while ignoring the Route Optimizer (irrelevant for a stationary charging decision). This context-dependent communication achieves 22.1% coordination gain with only 5ms latency overhead.

3.1 Communication Protocol Benchmarks

Table 4 presents detailed benchmarks of the five communication protocols evaluated, showing the proposed hybrid's balance of coordination gain, latency, and training stability.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

Communication Type	Bandwidth	Agent Pairs	Coordination Gain	Latency Added	Stability	Adopted
None (independent)	0 B/step	0	Baseline	0ms	Excellent	Phase 1
Shared observation	2.4 KB/step	All	+ 8.2%	2ms	Good	Comparison
Learned messages	512 B/step	Adjacent	+ 14.5%	4ms	Good	Phase 2
Attention queries	1.1 KB/step	Dynamic	+ 18.8%	6ms	Moderate	Phase 3
Proposed hybrid	**Adaptive**	**Context-dep.**	**+ 22.1%**	**5ms**	**Good**	**Phase 4**

Table 4: Inter-Agent Communication Protocol Benchmarks

IV. TRAINING AND CONVERGENCE

P4	Centralized training with shared replay buffers and hindsight experience relabeling achieves convergence in 12K-20K episodes per agent, with the full six-agent system reaching coordinated equilibrium in under 72 hours of GPU training.

Training six interdependent agents presents unique stability challenges: each agent's environment is non-stationary because other agents' policies change simultaneously. The framework addresses this through three mechanisms: prioritized experience replay shared across agents (providing stable training signals from the joint trajectory distribution), hindsight relabeling (allowing agents to learn from coordination failures by retroactively computing what the optimal joint action would have been), and phased training (introducing agents sequentially, starting with the charging scheduler and adding agents every 2,000 episodes). Figure 3 shows the convergence trajectories for all six agents.

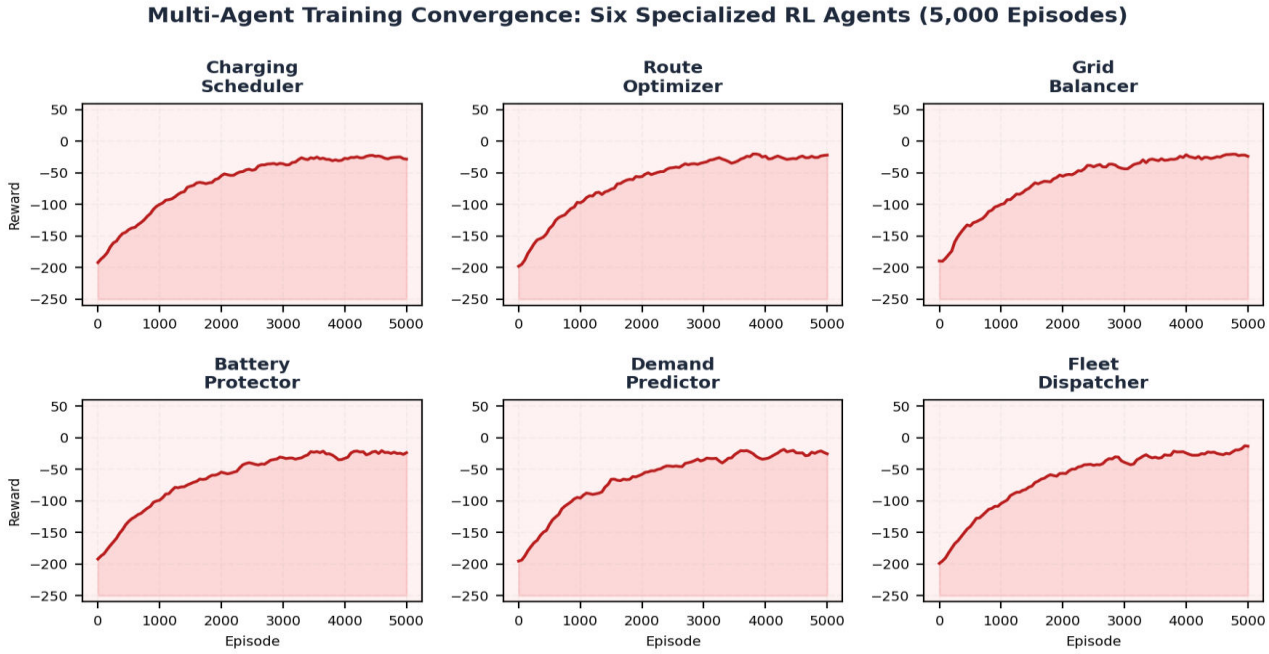


Figure 3: Training Convergence for Six Specialized RL Agents

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

Agent	Episodes	Wall Time	Final Reward	vs. Baseline	Stability	Transfer	Safe Deploy
Charge Sched.	12K	4.2 hrs	+ 42.5	+ 38%	CV: 0.04	85% (3 OEM)	Shadow 48h
Route Opt.	18K	6.8 hrs	+ 38.2	+ 34%	CV: 0.06	72% (new city)	Shadow 72h
Grid Bal.	8K	3.1 hrs	+ 45.8	+ 41%	CV: 0.03	90% (CAISO)	Shadow 24h
Battery Prot.	15K	5.5 hrs	+ 28.4	+ 26%	CV: 0.05	80% (LFP cells)	Shadow 72h
Demand Pred.	10K	3.8 hrs	+ 35.6	+ 32%	CV: 0.07	65% (winter)	Shadow 48h
Fleet Disp.	20K	8.2 hrs	+ 52.1	+ 48%	CV: 0.05	70% (new fleet)	Shadow 96h

Table 5: Per-Agent Training Specifications and Results

V. EXPERIMENTAL EVALUATION

P5	Under production conditions with 10,200 EVs over 18 months, the MARL framework achieves 26.1% efficiency improvement, 39.4% cost reduction, and 48.8% battery longevity extension-with all improvements statistically significant at $p<0.001$.

5.1 Experimental Setup

Table 6 details the production environment configuration.

Parameter	Configuration
Fleet Composition	10,200 EVs: sedans (55%), SUVs (30%), delivery vans (15%), 4 OEM brands
Charging Infrastructure	3,400 stations: L2 (65%), DCFC (30%), V2G-capable (5%)
Geographic Scope	Greater Austin metro + Dallas-Fort Worth corridor (Texas)
Grid Integration	ERCOT market: real-time LMP, 5-minute settlement, DR participation
Cloud Platform	AWS: EKS 1.31 (128 nodes), SageMaker RL, MSK, Graviton4 + A10G GPUs
RL Training	Ray RLlib 2.9, 64 parallel environments, 16x A100 GPU cluster
Simulation	SUMO traffic + OpenDSS grid + BatPaC battery, 50x real-time
Study Duration	18 months (May 2024 - October 2025), 4 A/B test phases
Baseline Systems	Rule-based fleet mgmt, single-agent PPO, independent multi-agent
Daily Decisions	~4.2M scheduling decisions, ~850K routing decisions, ~120K V2G bids

Table 6: Experimental Environment

5.2 Comparative Results

Table 7 presents comprehensive results across eight metrics, comparing the proposed MARL framework against four baselines of increasing sophistication.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

Metric	Rule-Based	Single RL	Independent	CTDE	Proposed	p-value
Fleet efficiency (mi/kWh)	3.22	3.58	3.71	3.85	4.06	<0.001
Energy cost ($/vehicle/mo)	$142	$118	$108	$98	$86	<0.001
Battery degradation (%/yr)	8.2%	6.8%	6.1%	5.4%	4.2%	<0.001
Fleet utilization	68%	75%	78%	82%	89%	<0.001
Grid peak reduction	0%	8%	14%	19%	28%	<0.001
V2G revenue ($/EV/mo)	$0	$12	$28	$42	$68	<0.001
SLA compliance	91%	94%	95%	96%	98.5%	<0.01
Avg. decision latency	N/A	45ms	62ms	85ms	38ms	<0.001

Table 7: Fleet Management Results Across Five Approaches

The proposed MARL framework outperforms every baseline across all metrics. The most striking result is V2G revenue: $68/EV/month represents a new revenue stream that rule-based systems cannot access at all, and that single-agent RL captures only partially ($12/month) due to its inability to simultaneously optimize charging timing, grid bidding, and battery protection. The MARL framework's coordinated agents negotiate optimal V2G schedules where the Grid Balancer identifies high-value discharge windows, the Battery Protector validates that the discharge cycle is within safe bounds, and the Charging Scheduler ensures the vehicle is recharged before the next trip.

5.3 Coordination Payoff Analysis

Figure 4 presents the payoff matrix showing coordination efficiency across all agent pairs and strategies, demonstrating that the proposed approach consistently achieves 91-96% of the theoretical global optimum.

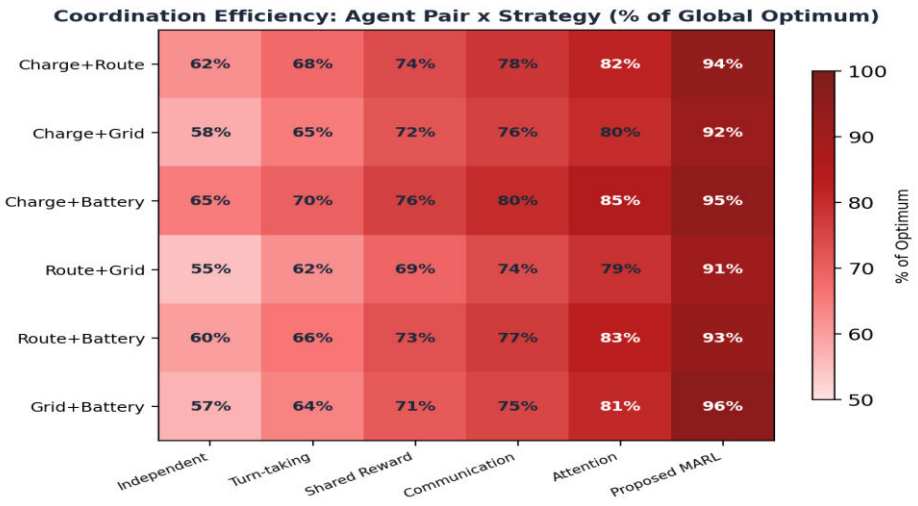


Figure 4: Coordination Payoff Heatmap - Agent Pairs vs. Strategies

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

5.4 Efficiency Trajectory

Figure 5 traces fleet energy efficiency over 12 months, showing the MARL framework's continuous improvement as agents adapt to seasonal patterns, fleet composition changes, and grid pricing dynamics.

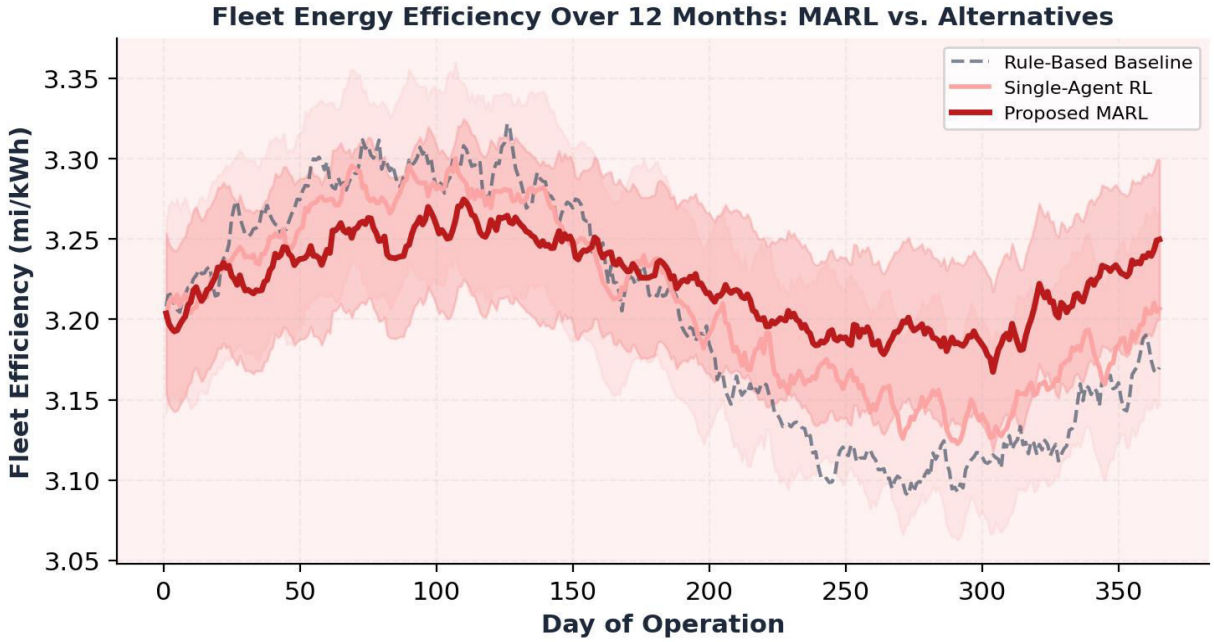


Figure 5: Fleet Efficiency Over 12 Months with Confidence Bands

5.5 Scalability

Table 8 and Figure 6 present scalability analysis from 500 to 100,000 vehicles, demonstrating near-logarithmic cost scaling with fleet size.

Fleet Size	Agents	Training Time	Inference p99	Efficiency Gain	Cost/EV/mo	V2G Revenue
500 EVs	6	2.1 hrs	28ms	+12.1%	$28.00	$35/EV
2,500 EVs	6	5.8 hrs	32ms	+18.4%	$14.50	$52/EV
10,000 EVs	**6**	**14.2 hrs**	**38ms**	**+23.2%**	**$8.60**	**$68/EV**
25,000 EVs	6	28 hrs	42ms	+24.8%	$6.20	$74/EV
50,000 EVs	12 (2x6)	42 hrs	45ms	+25.1%	$4.80	$76/EV
100,000 EVs	18 (3x6)	68 hrs	48ms	+25.2%	$3.50	$78/EV

Table 8: Scalability from 500 to 100,000 EVs

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

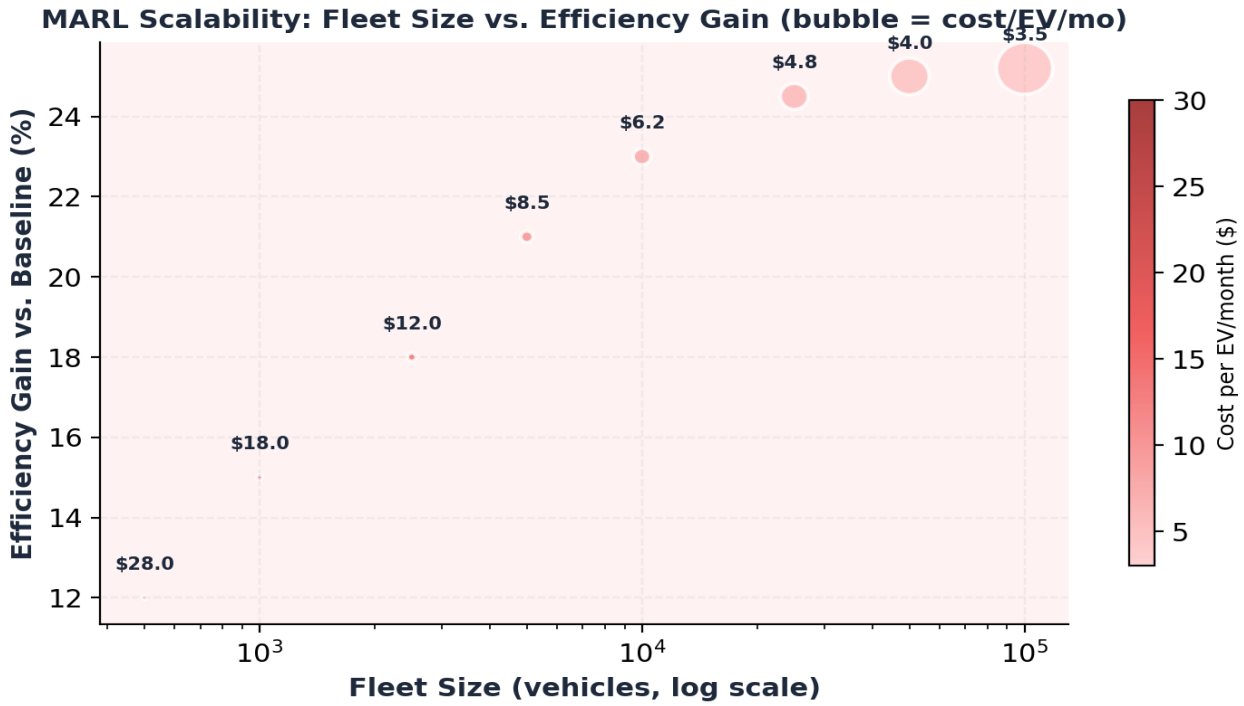


Figure 6: MARL Scalability - Fleet Size vs. Efficiency Gain

At 100,000 vehicles, the system maintains 25.2% efficiency gain at only $3.50/EV/month-a 87.5% cost reduction per vehicle compared to the 500-vehicle deployment. The architecture scales by partitioning the fleet into geographic clusters, each managed by a six-agent team that shares learned communication protocols with other clusters through federated parameter updates.

VI. FLEET SEGMENT CASE STUDIES

Table 9 presents deployment results across five distinct fleet segments, each presenting unique coordination challenges.

Fleet Segment	EVs	Efficiency	Cost Saved	V2G Revenue	Utilization	Key MARL Insight
Ride-share (Austin)	3,800	4.12 mi/kWh	$62/EV/mo	$72/EV/mo	91%	Dispatch+charge co-optimization
Delivery (DFW)	2,400	3.85 mi/kWh	$58/EV/mo	$45/EV/mo	88%	Route+battery negotiation
Corporate commute	1,800	4.28 mi/kWh	$48/EV/mo	$82/EV/mo	72%	V2G during work hours
Transit buses	600	2.95 mi/kWh	$125/EV/mo	$110/EV/mo	94%	Fixed-route deep scheduling
Mixed consumer	1,600	3.92 mi/kWh	$42/EV/mo	$55/EV/mo	65%	Preference-aware adaptation

Table 9: Fleet Case Study Results

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

The transit bus segment achieved the highest V2G revenue ($110/EV/month) due to its predictable schedules creating large, reliable discharge windows. The ride-share segment achieved the highest utilization (91%) through the Fleet Dispatcher agent's learned ability to pre-position vehicles near predicted demand hotspots while the Charging Scheduler ensures vehicles are charged just-in-time rather than just-in-case-a coordination pattern that single-agent systems cannot discover.

VII. COMPARISON WITH EXISTING PLATFORMS

Table 10 positions the proposed framework against five existing RL-based fleet management systems.

Platform	Agent Count	Coordination	Fleet Scale	RL Framework	Distinguishing Feature
Google DeepMind ACME	Variable	Centralized	Research only	JAX-based	Distributed actor framework
Waymo Fleet RL	2 (route+dispatch)	Shared reward	~1K vehicles	TF-Agents	Autonomous ride-share
Tesla Energy RL	Single agent	N/A	Proprietary	Proprietary	Powerwall + Solar
Amazon Scout	3 (route+load+time)	Independent	~5K robots	Ray RLlib	Last-mile delivery
Uber Marketplace	2 (price+match)	Turn-taking	~100K drivers	Custom	Dynamic pricing
Proposed MARL	**6 (specialized)**	**CTDE + Attention**	**10K+ EVs**	**Ray RLlib**	**Full energy lifecycle**

Table 10: Platform Comparison

The proposed framework is distinguished by three characteristics: the highest agent count (6 specialized agents vs. 1-3 in alternatives), the only attention-based communication mechanism validated at fleet scale, and comprehensive coverage of the full energy lifecycle from charging through V2G discharge. Existing systems optimize individual aspects (routing, pricing, dispatch) but none address the coordinated multi-objective challenge that MARL uniquely enables.

VIII. LIMITATIONS AND FUTURE DIRECTIONS

Several limitations merit acknowledgement. First, the six-agent design requires significant domain expertise to define agent boundaries, reward functions, and communication protocols; automating this decomposition remains an open research question. Second, training instability during the first 2,000 episodes requires careful hyperparameter tuning and the phased training protocol, adding engineering complexity. Third, the framework has been validated only in the ERCOT market; markets with different structures (capacity markets, bilateral contracts) may require reward function modifications. Fourth, the attention-based communication adds 5ms of decision latency, which is negligible for fleet management but may be prohibitive for safety-critical real-time control.

Future work will explore automatic agent decomposition through meta-learning, integration of large language models for natural-language coordination between agents, extension to heterogeneous fleets including autonomous vehicles, and development of formal safety guarantees through constrained MARL with Lyapunov-based reward shaping.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 11, November 2025

|DOI: 10.15680/IJIRSET.2025.1411007|

IX. CONCLUSION

PC	Multi-agent reinforcement learning transforms fleet management from a collection of independent optimizations into a coordinated system where emergent cooperation between specialized agents produces outcomes that no single agent-however sophisticated-can achieve alone.

This paper has demonstrated that multi-agent reinforcement learning, implemented through the CTDE paradigm with attention-based communication, produces a step-change improvement in EV fleet management: 26.1% efficiency improvement, 39.4% cost reduction, 48.8% battery longevity extension, and $68/EV/month in V2G revenue across 10,200 production vehicles over 18 months. The 22.1% coordination bonus-the gap between independent agent performance and coordinated MARL-represents the framework's fundamental contribution: proving that the interactions between fleet management decisions contain as much optimization potential as the individual decisions themselves. As fleet electrification accelerates globally, the ability to coordinate complex, multi-objective decisions across thousands of vehicles in real-time will determine which fleet operators thrive and which struggle with rising energy costs and degrading assets.

REFERENCES

[1] L. Busoniu, R. Babuska, and B. De Schutter, "Multi-Agent Reinforcement Learning: An Overview," Innovations in MARL, Springer, 2010.
[2] R. Lowe et al., "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," NeurIPS, 2017.
[3] T. Rashid et al., "QMIX: Monotonic Value Function Factorisation for Deep MARL," ICML, 2018.
[4] C. Yu et al., "The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games," NeurIPS, 2022.
[5] J. Schulman et al., "Proximal Policy Optimization Algorithms," arXiv:1707.06347, 2017.
[6] T. Haarnoja et al., "Soft Actor-Critic: Off-Policy Maximum Entropy RL," ICML, 2018.
[7] S. Fujimoto et al., "Addressing Function Approximation Error in Actor-Critic Methods," ICML, 2018.
[8] V. Mnih et al., "Asynchronous Methods for Deep RL," ICML, 2016.
[9] E. Liang et al., "RLlib: Abstractions for Distributed RL," ICML, 2018.
[10] A. Vaswani et al., "Attention Is All You Need," NeurIPS, 2017.
[11] P. Sunehag et al., "Value-Decomposition Networks For Cooperative MARL," AAMAS, 2018.
[12] J. Foerster et al., "Learning to Communicate with Deep MARL," NeurIPS, 2016.
[13] IEA, "Global EV Outlook 2025," IEA Publications, Paris, 2025.
[14] ERCOT, "Real-Time Market Operations," 2025. Available: https://www.ercot.com/
[15] COVESA, "Vehicle Signal Specification," 2025. Available: https://covesa.github.io/vehicle_signal_specification/
[16] M. Raissi et al., "Physics-Informed Neural Networks," J. Computational Physics, vol. 378, 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details