Joint V2I Optimization for Autonomous Driving: From Reinforcement Learning to LLMs and Quantum Decision-Making

Zijiang Yan, Hina Tabassum

Paper · Website · Demo

Contributions

- We develop an MORL framework to design joint network selection and autonomous driving policies in a multi-band vehicular network (VNet). The objectives are to
  - **i) maximize the traffic flow and minimize collisions by controlling the vehicle's motion dynamics (i.e., speed and acceleration) from a transportation perspective,** and
  - **ii) maximize the data rates and minimize handoffs (HOs) by jointly controlling the vehicle's motion dynamics and network selection from telecommunication perspective.**

  We consider a novel reward function that maximizes data rate and traffic flow, ensures traffic load balancing across the network, *penalizes HOs*, and unsafe driving behaviors.
- The considered problem is formulated as a **multi-objective Markov decision process (MOMDP)** that has **two-dimensional action space and rewards** consist of telecommunication and autonomous driving utilities. We then propose single policy MORL solutions with predefined preferences thus converting the MOOP into a single-objective and apply DQN and double DQN solutions. The resulting optimal policy depends on the relative preferences of the objectives.
- Learning optimized policies across **multiple preferences** remains challenging. To address this, we then develop a novel envelope MORL solution to effectively navigate the entire spectrum of preferences within a given domain. This approach empowers the trained model to generate the best possible policy tailored to any user-defined preference. Our algorithm hinges on two fundamental insights: firstly, we demonstrate that the **optimality operator** governing a generalized Bellman equation with preferences exhibits valid contraction properties. Secondly, by optimizing for the **convex envelope of multi-objective Q-values**, we ensure an efficient alignment between preferences and the resultant optimal policies. **Leveraging hindsight experience replay,** we recycle instructions to facilitate learning across various sampled preferences, while employing homotopy optimization to maintain manageable learning processes.

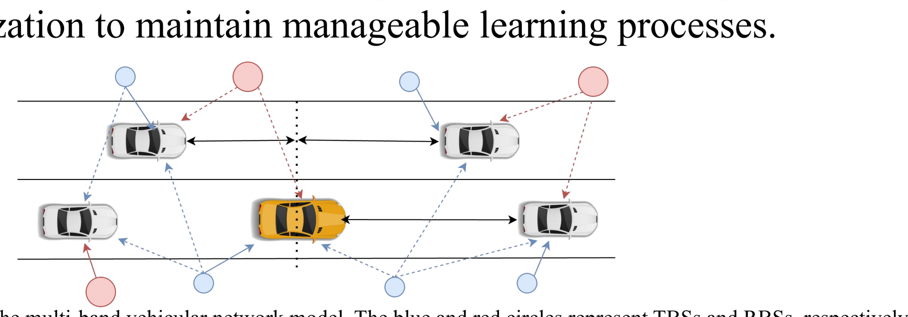

Figure 1: An illustrative structure of the multi-band vehicular network model. The blue and red circles represent TBSs and RBSs, respectively. The solid and dash line represent desired signal links and interference links, respectively.

System Model and Assumption

- **Kinematics Model:** $\frac{\partial}{\partial t}(x_j) = v_j \cos(\psi_j + \beta_j), \quad \beta_j = \arctan\left(\frac{\tan \delta_j^{\mathrm{fa}}}{2}\right)$
$$\frac{\partial}{\partial t}(y_j) = v_j \sin(\psi_j + \beta_j)$$
$$\frac{\partial}{\partial t}(v_j) = a_j, \quad \frac{\partial}{\partial t}(\psi_j) = \frac{v_j}{l_j}\sin\beta_j$$
- **Acceleration and Lane Change**
$$\frac{\partial}{\partial t}(\psi_j) = K_j^{\psi}\left[\left(\psi_{L_j} + \arcsin\left(\frac{\tilde{v}_{i,y}}{v_j}\right) - \psi_j\right]$$
$$a_j = K_0^{v}(v_r - v_j)$$
- **Network Composition:** two-tier downlink network with N_R RF BSs (RBSs) and N_T THz BSs (TBSs) supporting V (AVs) on a four-lane highway.
- **Bandwidth and Data Rate:** Each BS, whether RBS or TBS, is allocated a specific bandwidth (W_R or W_T), and data rates are computed as
$$R_{ij} = \frac{W_j}{\ln 2}\left[\ln(1+\mathrm{SINR}_{ij}) - \sqrt{\frac{V}{L_B}} f_Q^{-1}(\epsilon_c)\right] \qquad \mathrm{WR}_{ij} = \frac{R_{ij}}{\min(Q_i, n_i)}(1-\mu)$$
- **BS Quota and Selection:** Maximum AV limits for each RBS and TBS are denoted by Q_R and Q_T respectively. Each AV maintains a set of top three BSs based on data rates, provided $\mathrm{SINR}_{ij}(t) \geq \gamma_{th}$
- **Handoff Management:** AVs may switch BSs based on SINR requirements impacting data rates due to handoff (HO) latencies. A HO penalty μ is imposed to discourage frequent HOs, higher for TBSs and lower for RBSs.

MOMDP Formulation

- **State Space:** position, velocity, number of AVs associated with BS i, and their respective SINRs with BSs.
$$\mathcal{S} = \begin{bmatrix} x_1 & y_1 & v_1 & \psi_1 & n_R^1 & n_T^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{M_1} & y_{M_1} & v_{M_1} & \psi_{M_1} & n_R^{M_1} & n_T^{M_1} \end{bmatrix}$$
- **2D Action Space:** lane changes, acceleration, stop, and deceleration. Communication Action includes different strategies for selecting BS.
$$\mathcal{A} = \begin{bmatrix} \{a_{\mathrm{tele}}^1, a_{\mathrm{tran}}^1\} & \{a_{\mathrm{tele}}^1, a_{\mathrm{tran}}^2\} & \dots & \{a_{\mathrm{tele}}^1, a_{\mathrm{tran}}^5\} \\ \vdots & \vdots & \vdots & \vdots \\ \{a_{\mathrm{tele}}^3, a_{\mathrm{tran}}^1\} & \{a_{\mathrm{tele}}^3, a_{\mathrm{tran}}^2\} & \dots & \{a_{\mathrm{tele}}^3, a_{\mathrm{tran}}^5\} \end{bmatrix}$$
- **Reward Functions:**
$$r_t^{\mathrm{j,tran}} = c_1\left(\frac{v_t^j - v_{\min}}{v_{\max} - v_{\min}}\right) - c_2\cdot\delta_2 + c_3\cdot\delta_3 + c_4\cdot\delta_4,$$
$$r_t^{\mathrm{j,tele}} = c_5 \mathrm{WR}_{i^*,j,t}\left(1 - \min(1, \xi_t^j)\right)$$
$$\mathbf{Q}_\pi(s,a,\boldsymbol{\omega}) = \mathbb{E}_\pi\left[\sum_{j=1}^{M_1} r_t^{\mathrm{j,tran}}, \sum_{j=1}^{M_1} r_t^{\mathrm{j,tele}}\right]$$
where δ_2 is collision factor , ξ_t^j is HO probability

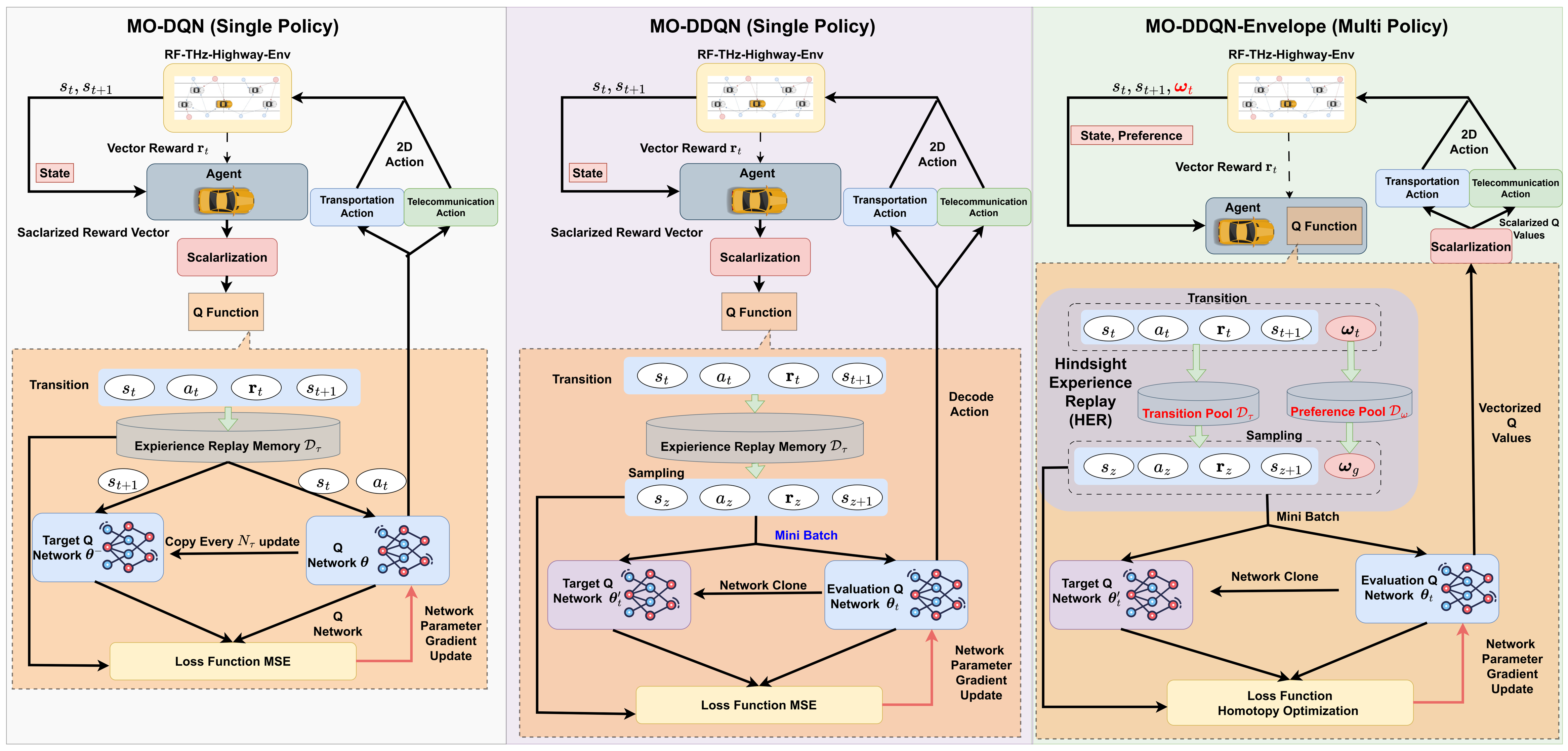

Figure 2: Comparison of MO-DQN, MO-DDQN, and the proposed MO-DDQN-envelope framework

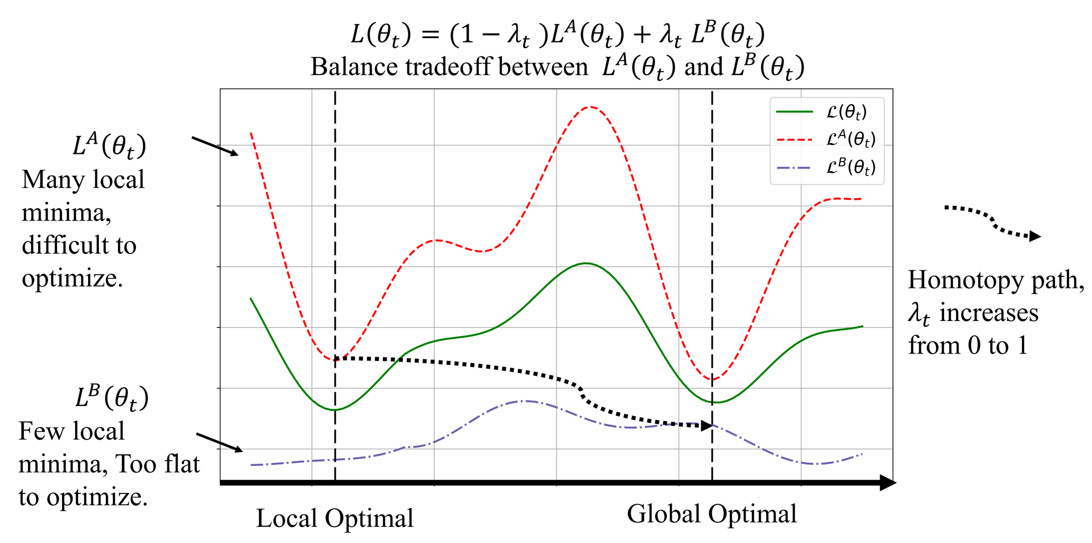

Figure 3: An explanation for homotopy optimization method used in the envelope deep MORL algorithm. The MSE loss $L^A(\theta)$ is hard for optimization since there are many local minima over its landscape. Although the value metric loss $L^B(\theta)$ has fewer local minima, it is also hard for optimization since there are many vectors Q minimizing value metric d. The landscape iof $L^B(\theta)$ is too flat. The homotopy path connecting $L^A(\theta)$ and $L^B(\theta)$ provides better opportunities to find the global optimal parameters θ∗

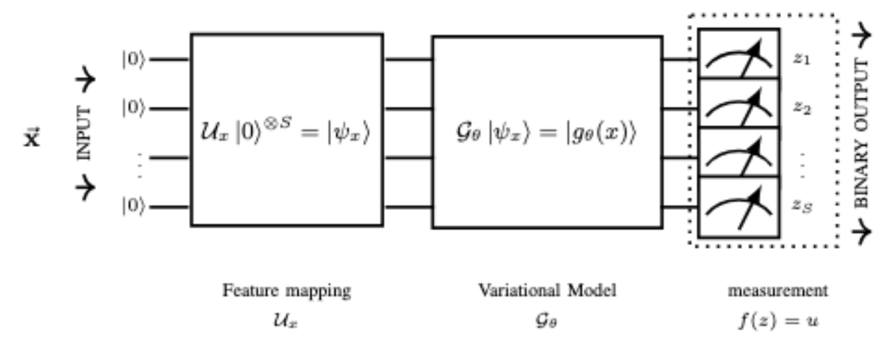

Figure 4: Overview of the Quantum Neural Network (QNN)

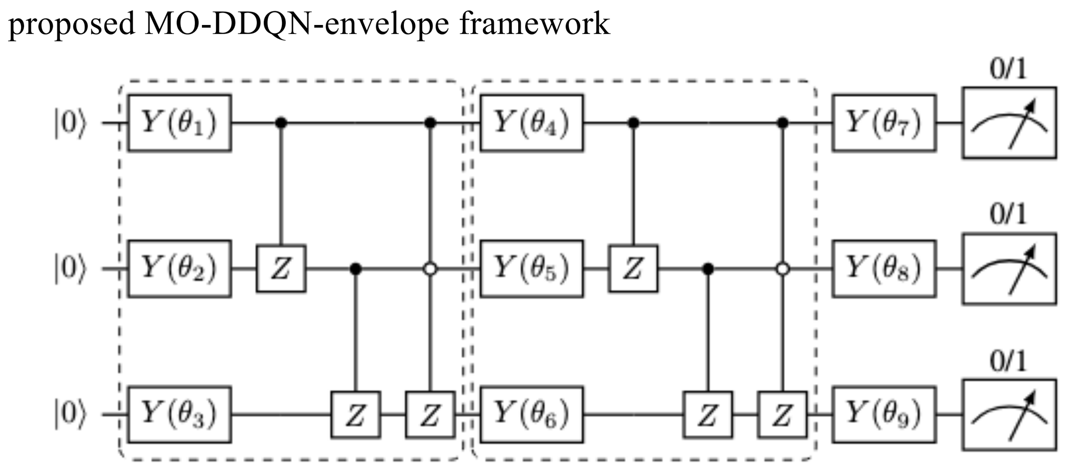

Figure 7: Overview of the Variation Quantum Circuit (Ansatz)

Simulation Results and Evaluation

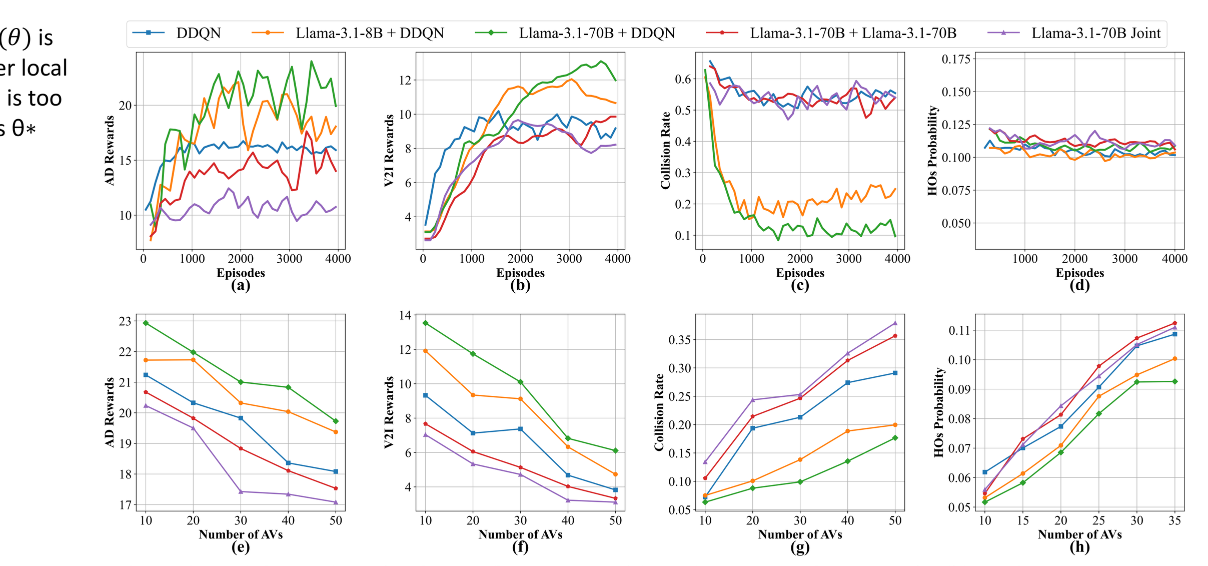

Figure 8: Training and Evaluation performance on (a) total transportation rewards, (b) total telecommunication rewards, (c) collision rate, and (d) HOs probability on Hybrid LLM-DDQN approach.

Proposed Hybrid LLM-DRL Solution

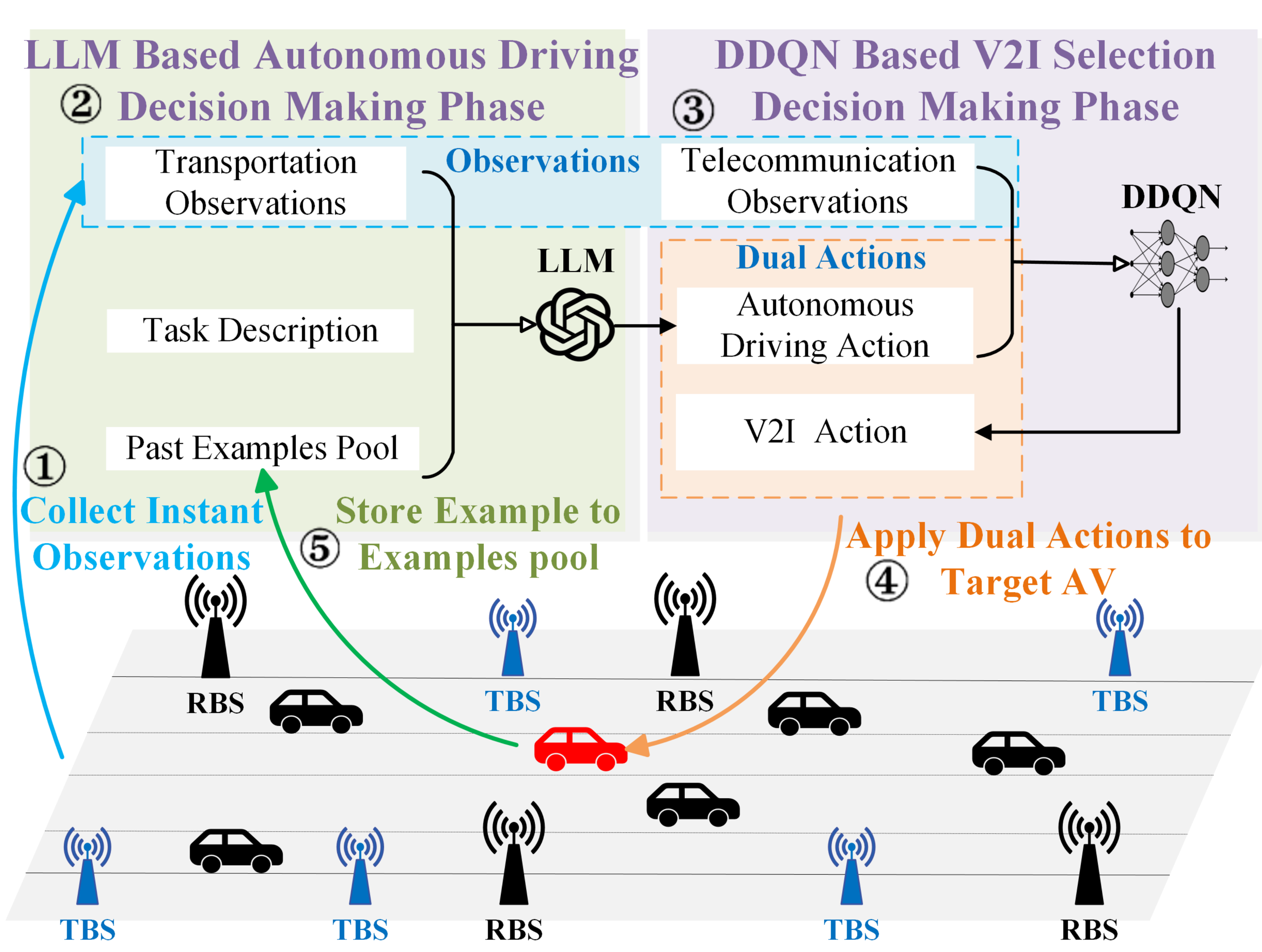

Figure 5: Hybrid LLM-DDQN Framework in the RF-THz-Highway Environment

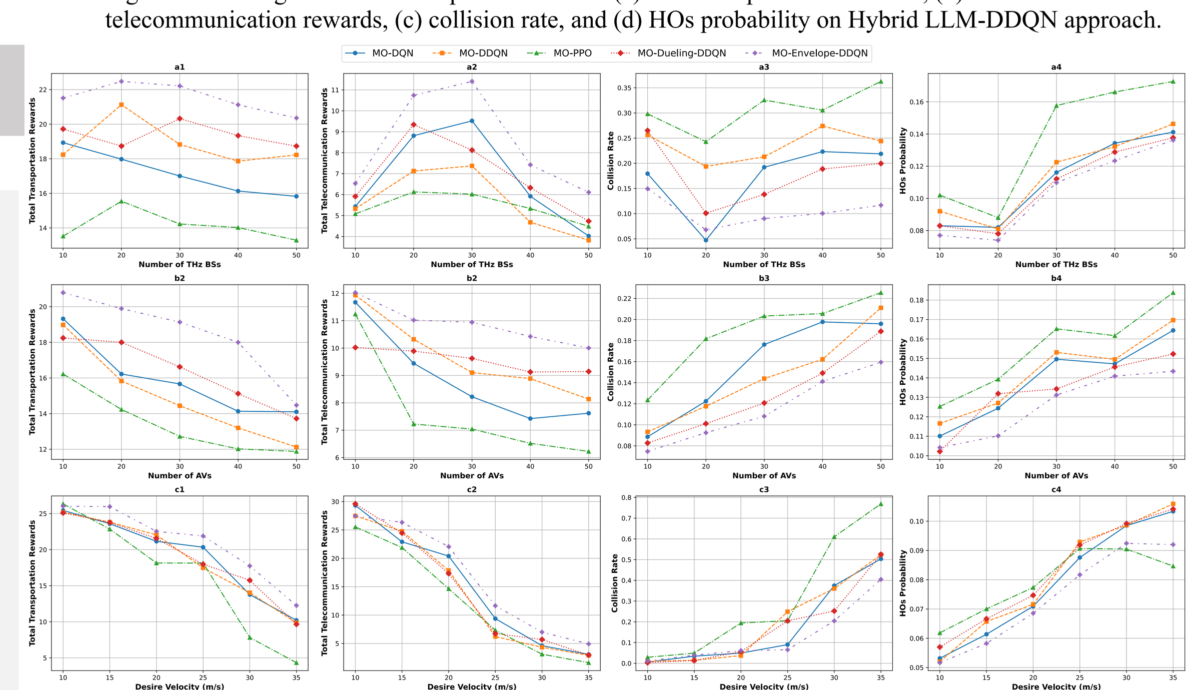

Figure **9**: Multi Objective Reinforcement Learning Evaluation performance on (1) total transportation rewards, (2) total telecommunication rewards, (3) collision rate, and (4) HOs probability, as a function of (a) variation in TBSs , (b) Variation in different number of AVs, (c) different desire speeds

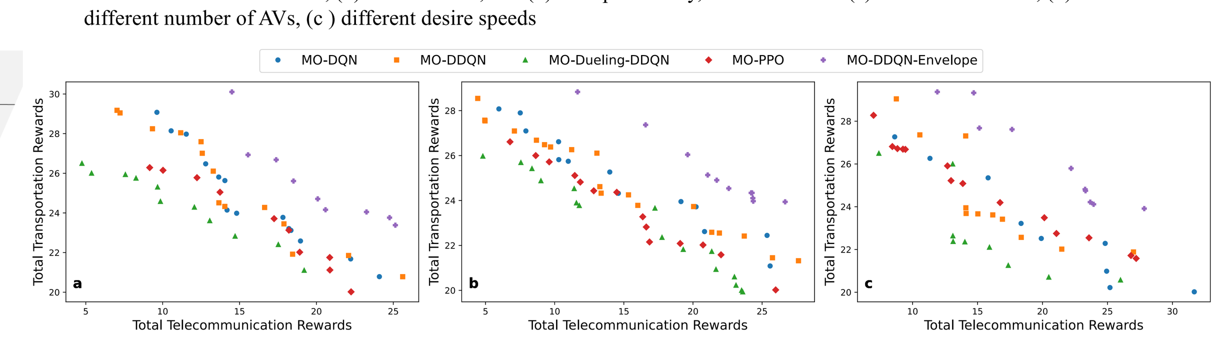

Figure **10**: **Muti objective Reinforcement Learning Approach** Pareto Frontier Comparison in MOO for total Transportation reward and total telecommunication reward among MO-DQN, MO-DDQN, MO-dueling-DDQN, MO-PPO, and MO-DDQN-Envelop, across instances:(a) I-(20,30,20,20), (b) I-(20,30,10,20), (c) I-(20,30,20,50)

Proposed Quantum Deep Learning Solution

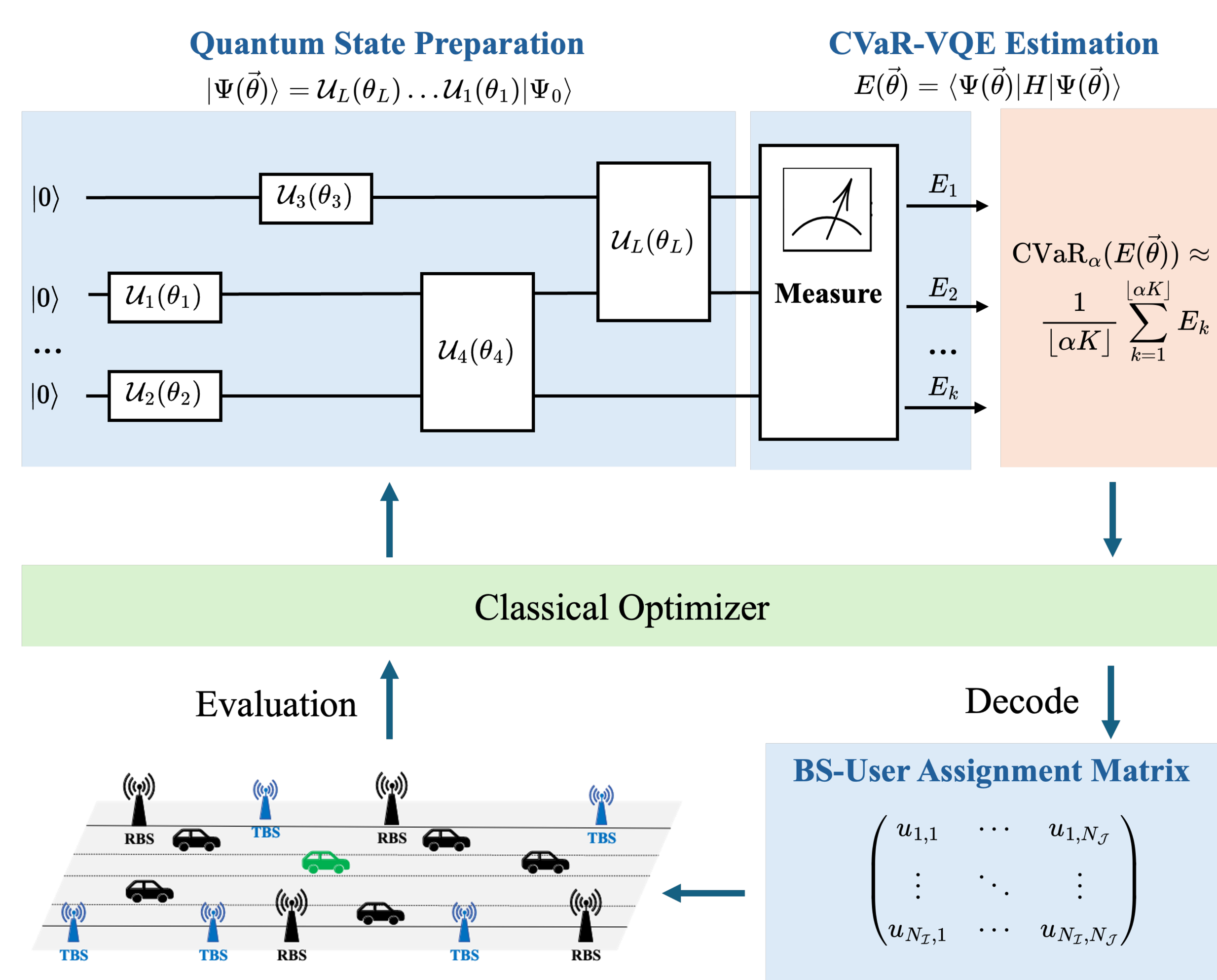

Figure 6 :Conditional Value at Risk (CVaR)-based VQE Framework with proposed Vehicular Network Model

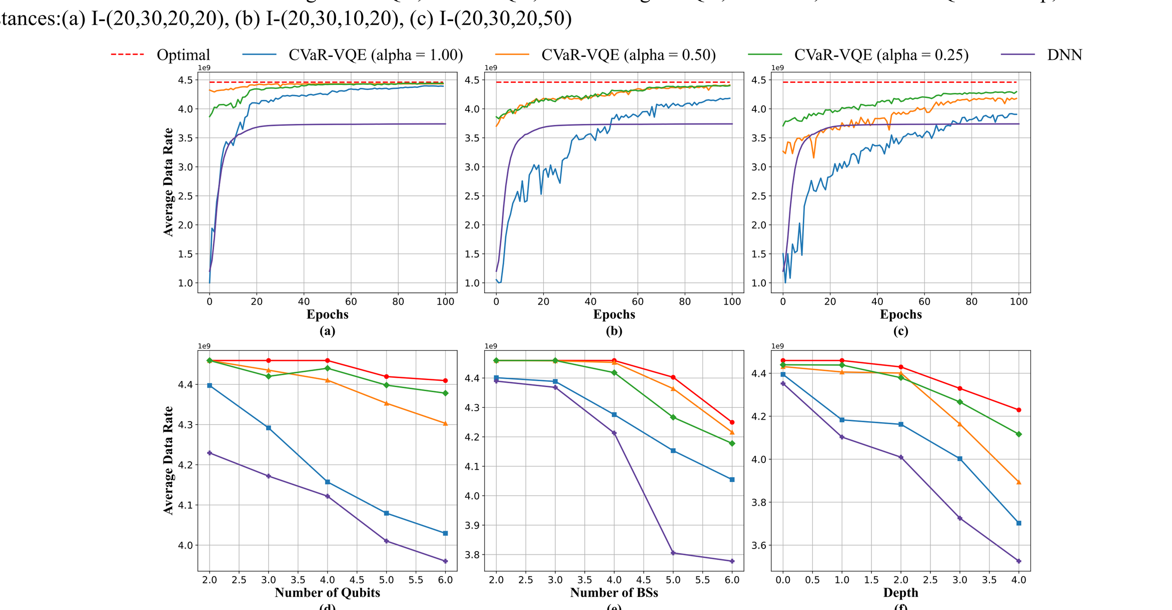

Figure 11: CVaR-VQE (Quantum neural network solution) on VNet Training and Evaluation Performance

References

- **Z. Yan** and **H. Tabassum**, "Generalized multi-objective reinforcement learning with envelope updates in URLLC-enabled vehicular networks," arXiv preprint arXiv:2405.11331, 2024.
- **Z. Yan** and **H. Tabassum**, "Reinforcement learning for joint V2I network selection and autonomous driving policies," in Proc. IEEE Global Commun. Conf. (GLOBECOM), Rio de Janeiro, Brazil, Dec. 2022, pp. 1241–1246.
- **Z. Yan**, H. Zhou, **H. Tabassum** and X. Liu, "Hybrid LLM-DDQN-Based Joint Optimization of V2I Communication and Autonomous Driving," in *IEEE Wireless Communications Letters*, vol. 14, no. 4, pp. 1214-1218, April 2025, doi: 10.1109/LWC.2025.3539638.
- **Z. Yan**, H. Zhou, J. Pei, A. Kaushik, **H. Tabassum**, and P. Wang, "CVaR-based variational quantum optimization for user association in handoff-aware vehicular networks," in Proc. IEEE Int. Conf. Commun. (ICC), Montreal, QC, Canada, Jun. 2025, accepted.

Acknowledgement