Optimizing Vehicular Networks with Variational Quantum Circuits-based Reinforcement Learning

Zijiang Yan, Ramsundar Tanikella, Hina Tabassum

IEEE INFOCOM 2024

YORK UNIVERSITÉ UNIVERSITY

Contributions

- We explore the integration of **variational quantum circuits (VQCs)** and reinforcement learning (RL) to optimize the kinematics and network connectivity of autonomous vehicles (AVs) in dynamic wireless and road-traffic flow environments.
- We employ VQC-based multi-objective RL to manage both **cell-association** and **autonomous driving policies** on a multi-lane highway. This includes BSs operating across RF and THz spectrums.
- We formulate the problem as a multi-objective **Markov decision process (MOMDP),** and transform this into **quantum eigen-states and eigen-actions** using quantum circuits.

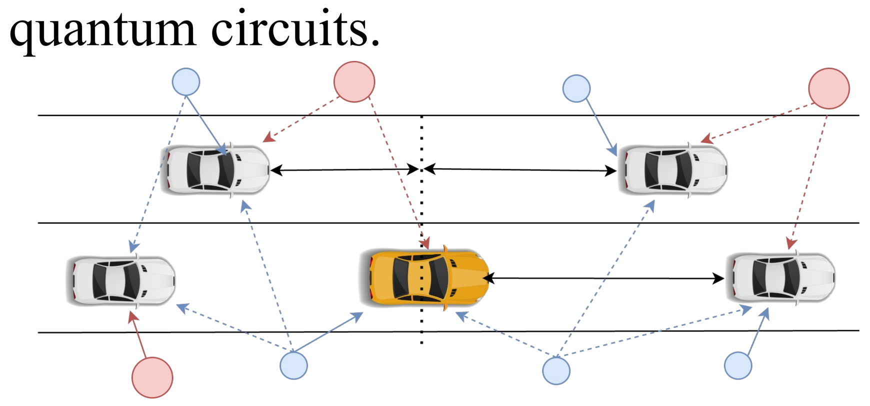

Figure 1: An illustrative structure of the multi-band vehicular network model. The blue and red circles represent TBSs and RBSs, respectively. The solid and dash line represent desired signal links and interference links, respectively.

System Model and Assumption

- **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 $R_{ij} = W_j \log_2\left(1 + \mathrm{SINR}_{ij}\right)$
- **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.
- **2D Action Space:** lane changes, acceleration, stop, and deceleration. Communication Action includes different strategies for selecting BS.
- **Reward Functions:**

$$r_j^{\text{tran}}(t) = \omega_1 \left(\frac{||\mathbf{v}_j(t)|| - v_{\min}}{v_{\max} - v_{\min}} \right) - \omega_2 \cdot \delta, \forall k \in \mathcal{U},$$

$$r_j^{\text{tele}}(t) = \omega_3 \hat{R}_{i_0 k}(t)\left(1 - \min(1, \xi_k(t))\right)$$

where δ is collision factor , ξ_t^j is HO probability

Proposed VQC-MORL Solution

- **VQC function approximator:** The VQC approximates the Q-function crucial for determining optimal actions, as given by:

$$Q(s,a;\theta) = \langle O_a \rangle_{s,\theta} = \langle 0^{\otimes 5} | U_\theta^\dagger(s) O_a U_\theta(s) | 0^{\otimes 5} \rangle$$

where $\langle O_a \rangle_{s,\theta}$ is the expectation of observables at the VQC output

- **Parameterization and Observables :** The VQC is parametrized by θ and adjusted so that the expected values of observables, O_a, fall within the real numbers, $E(O_a) \in R$
- **Loss Function:** Updated Q-values are incorporated into a loss function derived from Q-learning:

$$\mathcal{L}(\theta) = \frac{1}{|D|} \sum_{(s,a,r,s') \in D} \left(Q(s,a;\theta) - [r + \max_{a'} Q(s',a';\theta')]\right)^2$$

This loss function is used in a gradient descent step to optimize θ, improving the selection of action combinations for given states.

Algorithm 1: VQC-MORL Algorithm

Result: Quantum Circuit U_θ
Data: Quantum Circuit U_θ, Experience replay memory D, mini batch-size m

```
Initialization: D ← 0, θ ← 0, Target quantum circuits θ* ← θ, RBSs, TBSs, AVs
while episode < episode limit do
    t ← 0, s_1 initial and encode it to quantum state
    while t ≤ horizon limit do
        AV selects a_t by ε-greedy search as a_t^tele and a_t^tran and Enforce a_t^tele and a_t^tran to AV;
        Experience Replay: sample mini-batch transitions in D (s_k, a_k, r_k, s'_k) where k ∈ m;
        Set target-Q function: Q(s, a; θ) = ⟨O_a⟩_{s,θ}
        Set real Q-function: Q(s_t, a_t; θ)
        Compute loss: L(θ)
        Perform gradient descent step by minimizing loss L(θ); θ ← θ − a_t · L(θ) ·_θ y_k;
        Update the U_θ weights θ ← θ*;
    end
    Policy updated in terms of U_θ
end
```

Variation Quantum Circuit (VQC)

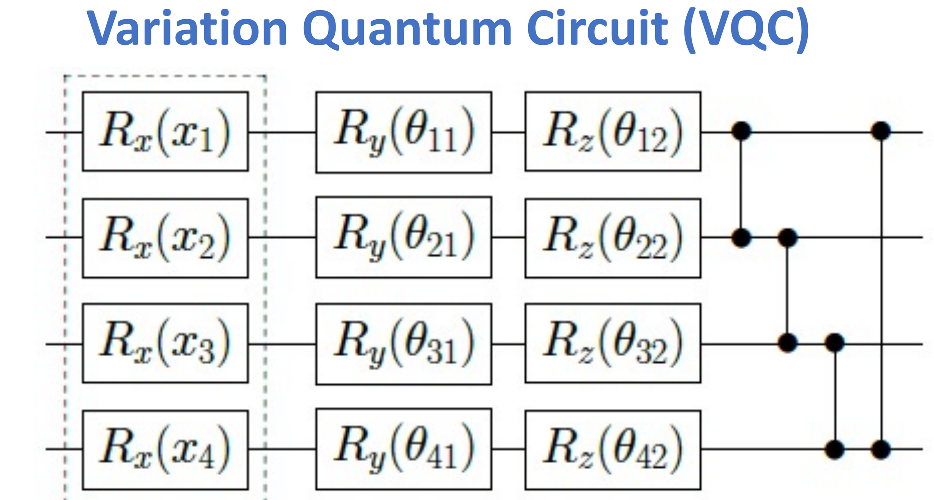

Figure 2: Skolik's Architecture: when data re-uploading is used, the whole circuit is repeated in each layer. Otherwise, just the part that is not surrounded by dashed lines.

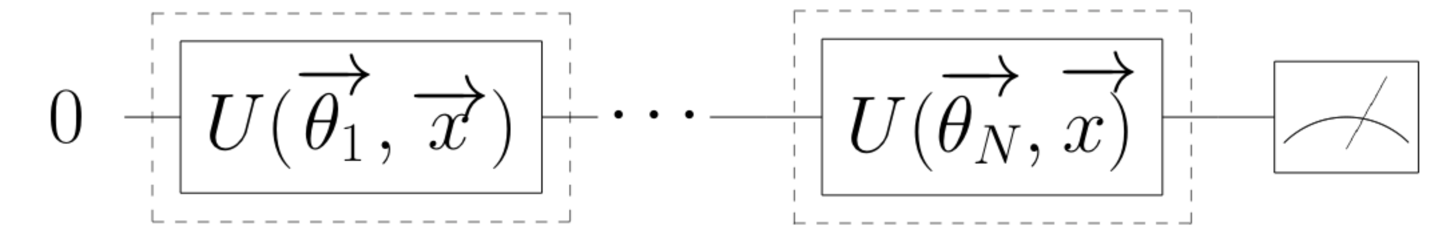

Figure 3: UQC Architecture. Each processing layer $\boldsymbol{U}$ is given by $U^{UAT}(\vec{x},\vec{\omega},\alpha,\varphi) = R_y(2\varphi)R_z(2\vec{\omega}\cdot\vec{x} + 2\alpha)$ and $\vec{\theta_i} = (\vec{\omega},\alpha,\varphi)$, Although a single-qubit ansatz was shown for simplicity, this ansatz can be generalized to allow multiple qubits.

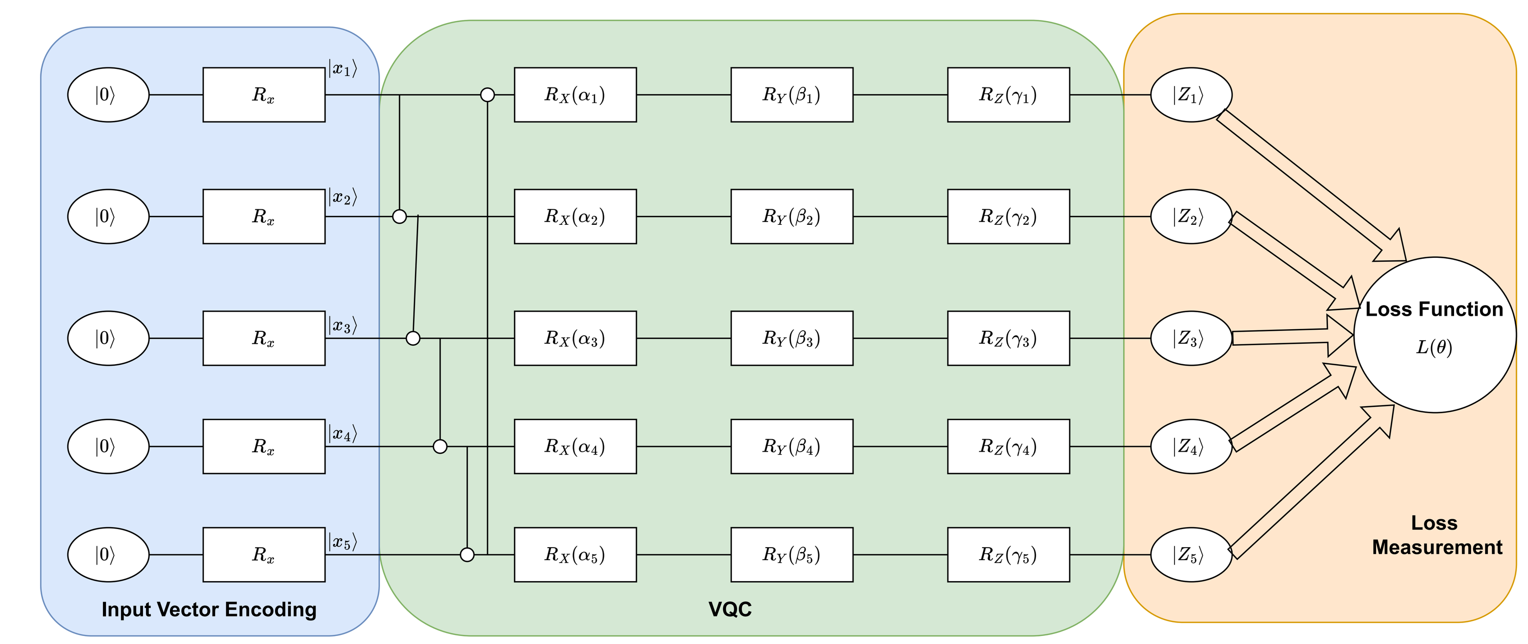

Figure 4: An illustration of the proposed VQC-MORL architecture where we have 5 qubits and 3 layers in the experiment. $\boldsymbol{R_x}$ and $\boldsymbol{R_z}$ are utilized for state encoding. 5 layers are repeated to approximate Q-function. The value function using 1-qubit Pauli-Z observable

Simulation Results and Evaluation

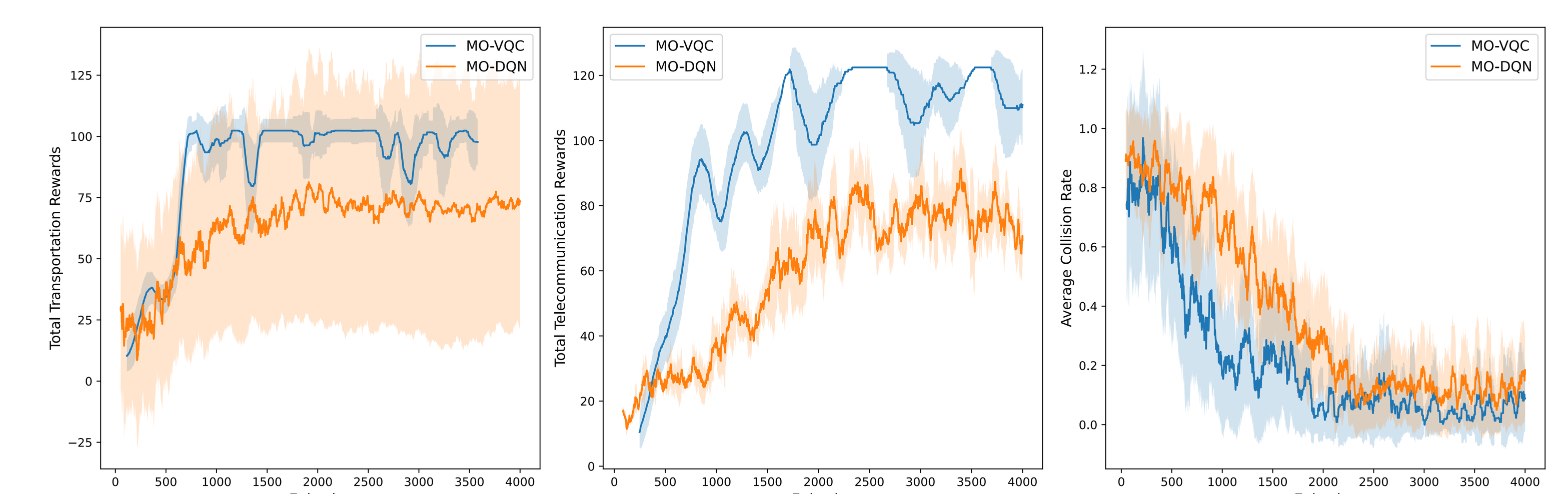

Figure 5: Training performances (ego vehicle): (a) Total telecommunication reward (b) Total transport reward (c) Collision Rate

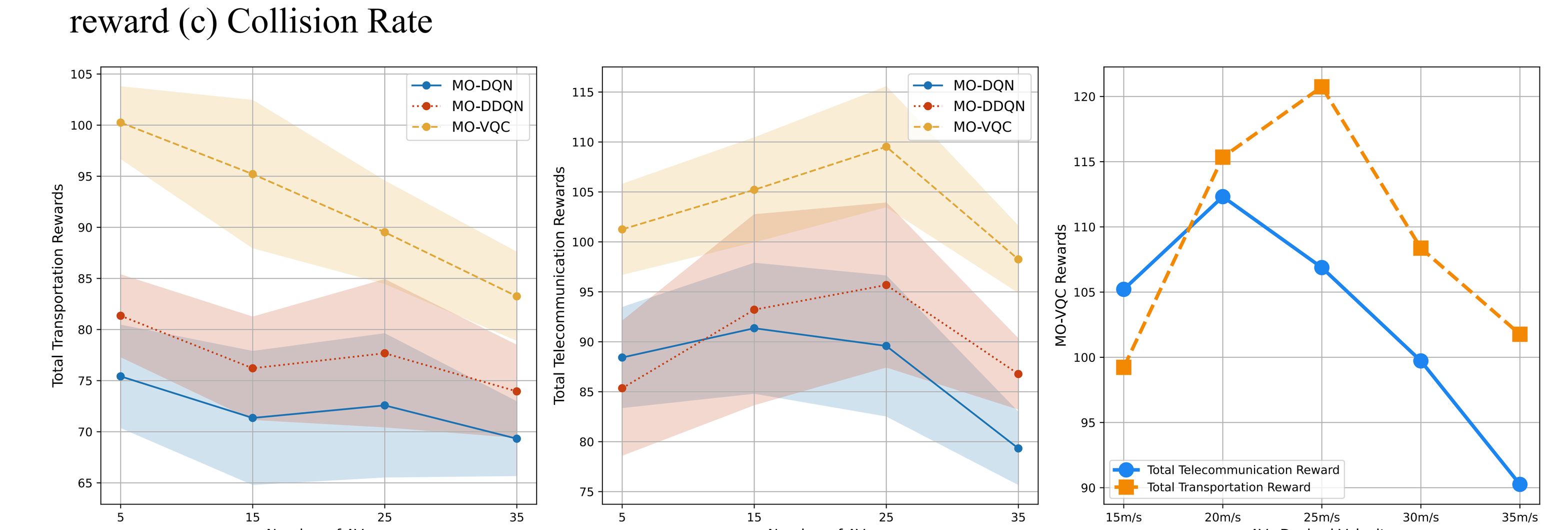

Figure 6: Evaluation performance (ego vehicle): (a) Total telecommunication reward (b) Total transport reward (c) Total reward. The considered VQC architecture has 5 qubits and 3 layers.

References

- Z. Yan, R. Tanikella, and H. Tabassum, *Optimizing Vehicular Networks with Variational Quantum Circuits-based Reinforcement Learning*, in IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), May 2024.
- Z. Yan and H. Tabassum, *Reinforcement learning for joint V2I network selection and autonomous driving policies*, in IEEE Global Communications Conference (pp. 1241-1246), Dec. 2022.
- R. Coelho, A. Sequeira, and P. L. Santos, *VQC-Based Reinforcement Learning with Data Re-uploading: Performance and Trainability*, arXiv preprint arXiv:2401.11555, 2024.

Acknowledgement

LASSONDE SCHOOL OF ENGINEERING | YORK U | mitacs