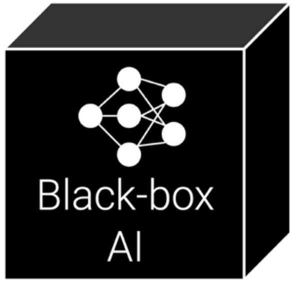
High-order Interactions Modeling for Interpretable Multi-Agent Q-Learning

Qinyu Xu Yuanyang Zhu† Xuefei Wu† Chunlin Chen

November 2025

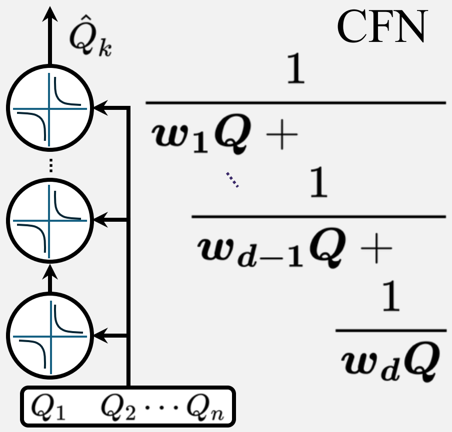
Why QCoFr?

- **Black-box models**
  - ➢ opaque value decomposition
- **Post-hoc explainable methods**
  - ➢ limited insight



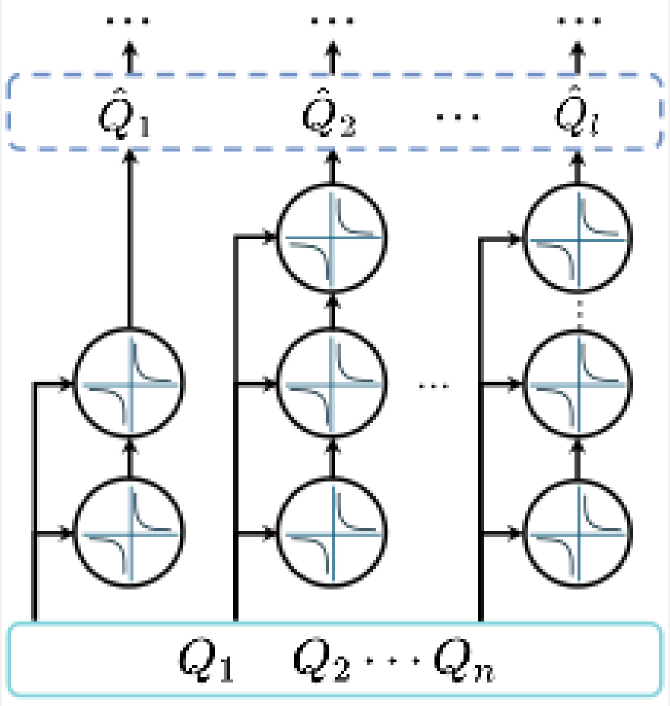
- **Combinatorial explosion**
  - ➢ hard to model high-order interactions

- **Continued fraction** structure
  - ✓ Intrinsic interpretable
  - ✓ Flexibly model arbitrary-order interactions
  - ✓ $\mathcal{O}(n)$ complexity



How CFN Captures High-order Interactions?

$$Q_{tot} = \sum_{k=1}^{l} \alpha_k \cdot \widetilde{f_k}(\boldsymbol{Q})$$



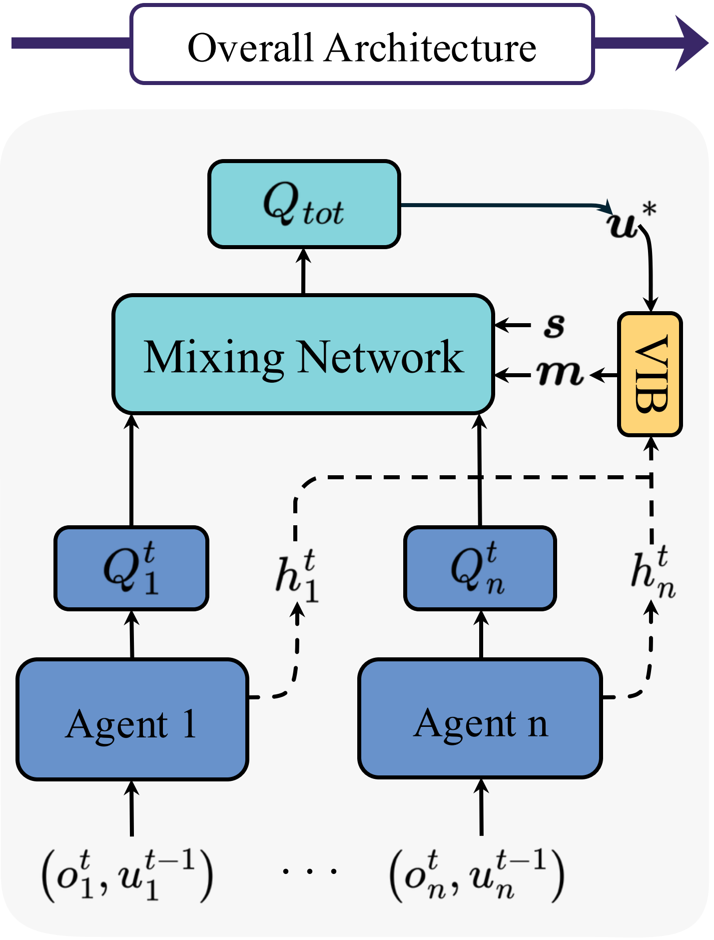
$$\widetilde{f_k}(\boldsymbol{Q}) = \cfrac{1}{\boldsymbol{w_1 Q} + \cfrac{1}{\boldsymbol{w_2 Q} + \cdots}} = \sum_{p_1,\ldots,p_n=0}^{\infty} c_{p_1,\ldots,p_n} \prod_{i=1}^{n} Q_i^{p_i}$$

$$f(\boldsymbol{Q}) - R_d(\boldsymbol{Q}) = \mathcal{O}\left(\boldsymbol{Q^{d+1}}\right)$$

$$\widehat{Q_k} = \cfrac{1}{\boldsymbol{w_1 Q} + \cfrac{1}{\boldsymbol{w_2 Q} + \cdots \cfrac{1}{\boldsymbol{w_d Q}}}} = \sum_{p_1,\ldots,p_n=0}^{d} c_{p_1,\ldots,p_n} \prod_{i=1}^{n} Q_i^{p_i}$$

- $c_{p_1,\ldots,p_n} \leftrightarrow \boldsymbol{w_k}$ (one-to-one correspondence)
- l ladders, each models interactions up to depth d
- $\mathcal{O}(n)$ complexity

Methods



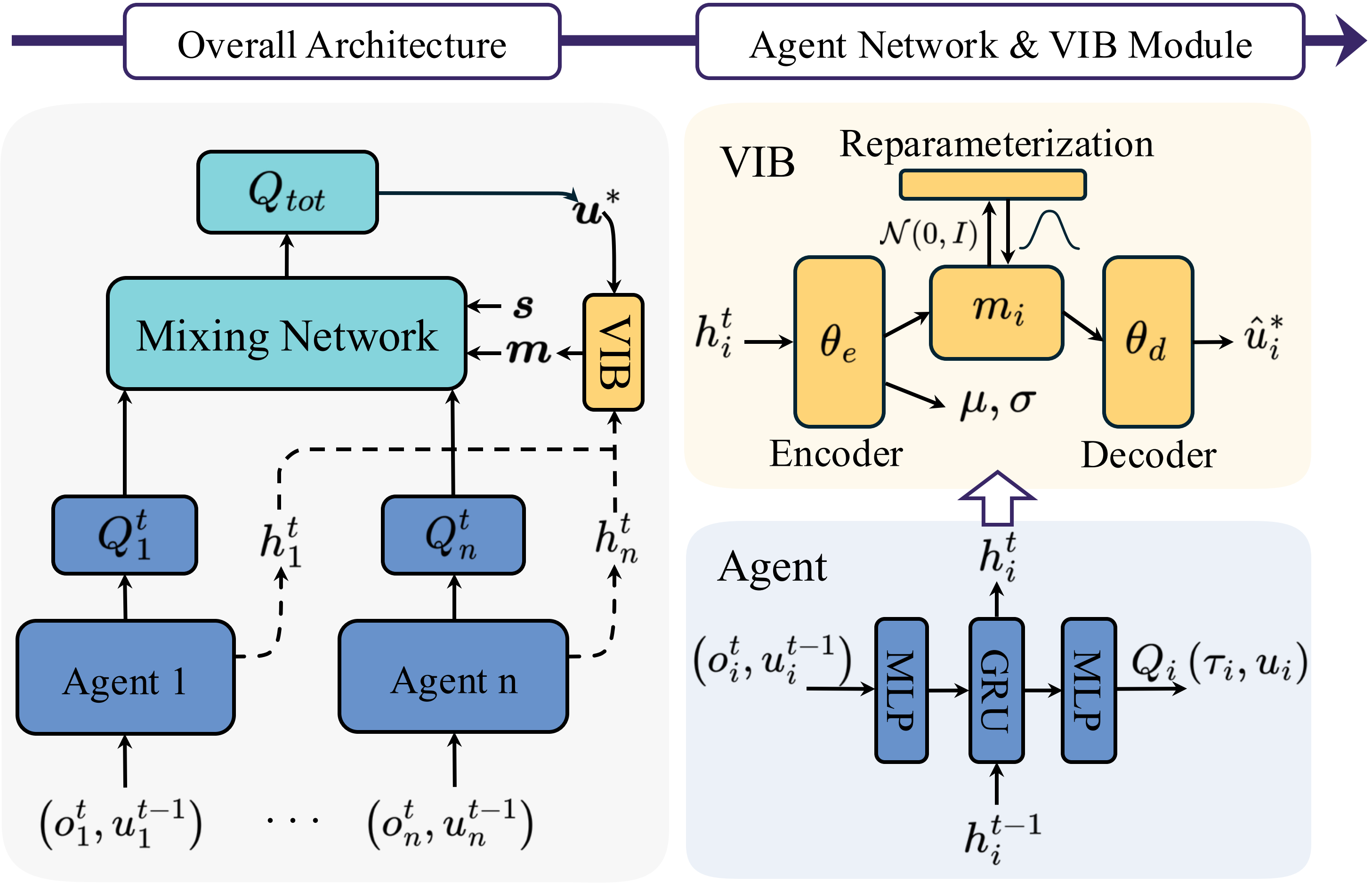
Overall Architecture

(1) Individual Action-value Function (**Agent Network**)

(2) Assistive Information Generation Module (**VIB Module**)

(3) Joint Action-value Function Q_{tot} (**CFN Mixer**)

VIB Module



From Markov Dependency:

$$\boldsymbol{h} \to \boldsymbol{m} \to \boldsymbol{u}^*$$

Optimization Objective:

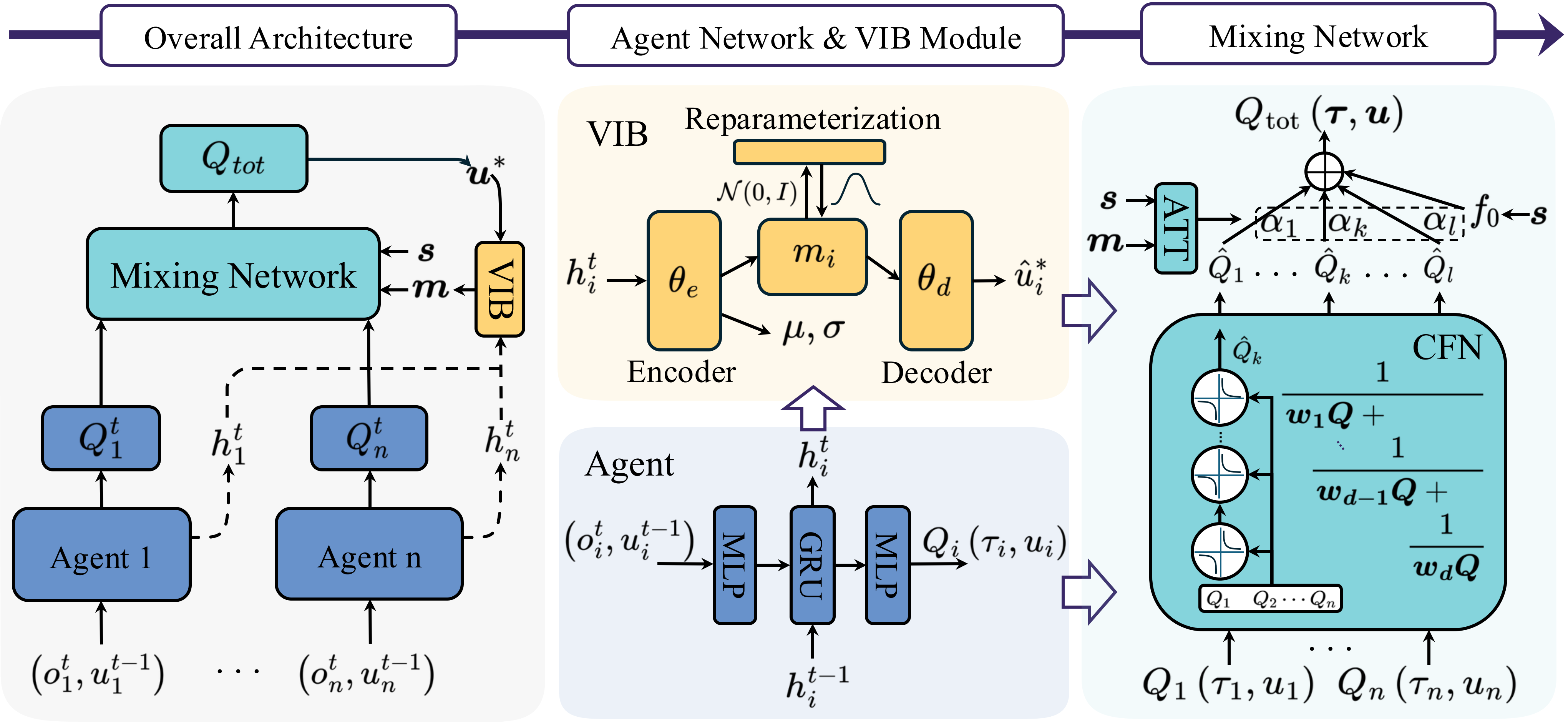
$$J_{IB}(\boldsymbol{\phi}) = I(\boldsymbol{m}, \boldsymbol{u}^*; \boldsymbol{\phi}) - \beta I(\boldsymbol{m}, \boldsymbol{h}; \boldsymbol{\phi})$$

Lower Bound for $I(\boldsymbol{m}, \boldsymbol{u}^*; \boldsymbol{\phi})$

Upper Bound for $I(\boldsymbol{m}, \boldsymbol{h}; \boldsymbol{\phi})$

$$\mathcal{L}_{VIB} = \frac{1}{N}\sum_{i=1}^{N} E_{\epsilon \sim p(\epsilon)}[-\log q\left(u_i^* \mid f(h_i, \epsilon)\right)] + \beta \mathrm{KL}[p(\boldsymbol{m} \mid h_i), \tilde{q}(\boldsymbol{m})].$$

CFN Mixer



- d-order interactions
- Non-negative activation function

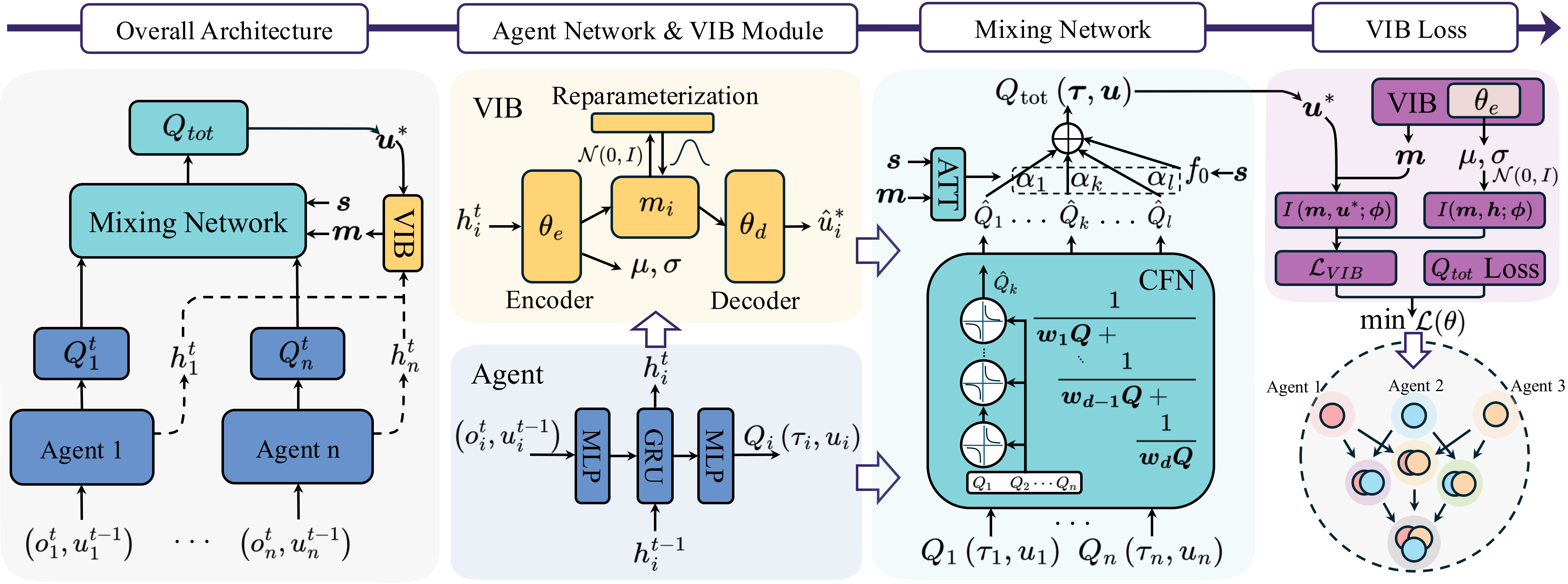
$$\frac{1}{\max(|z|, \delta)}$$

a small positive constant

- Extendable to non-IGM

$$Q_{tot} = \sum_{k=1}^{l} \alpha_k \widehat{Q_k} = \sum_{k=1}^{l} \alpha_k \sum_{p_1,\ldots p_n=0}^{d} c_{p_1,\ldots,\ldots p_n} \prod_{i=1}^{n} Q_i^{p_i} = \sum_{p_1\ldots p_n=0}^{d} c'_{p_1,\ldots,\ldots p_n} \prod_{i=1}^{n} Q_i^{p_i}$$

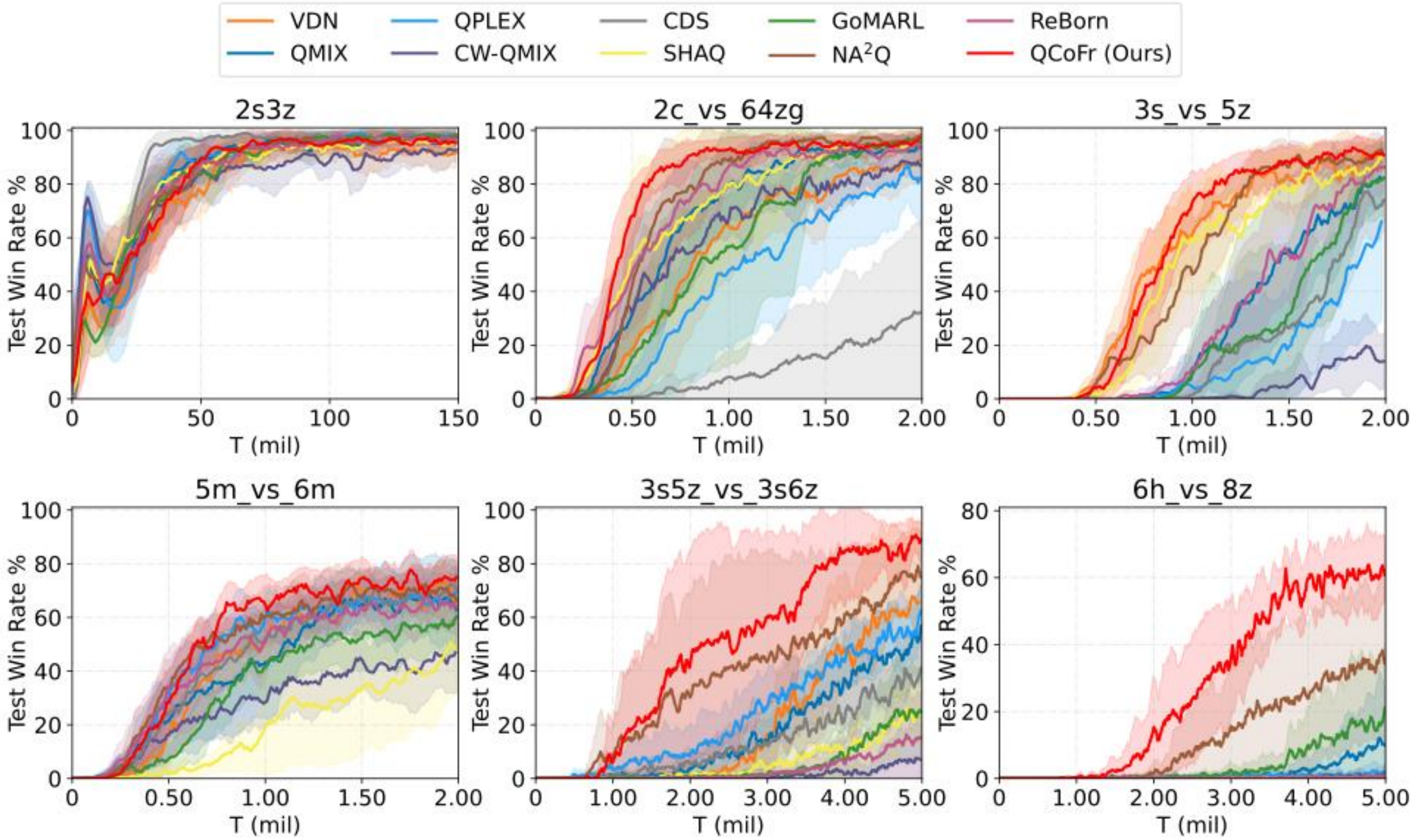
Overview



Overall Learning Objective: $\mathcal{L}(\theta) = \sum_i (Q_{tot}(s, \boldsymbol{\tau}, \boldsymbol{u}) - y_i)^2 + \mathcal{L}_{VIB}$

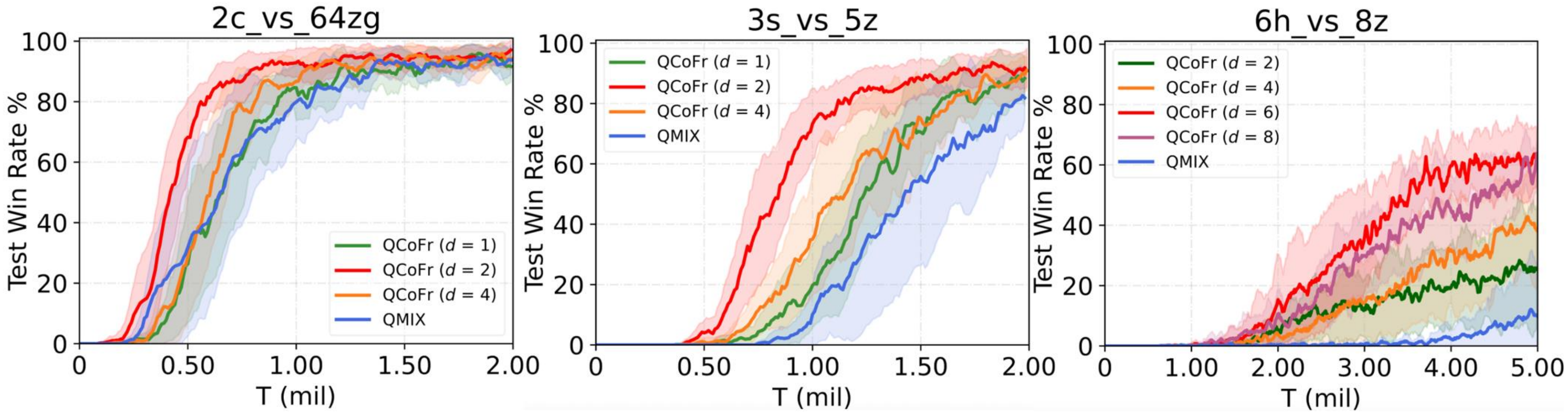
Experiments — SMAC

- **QCoFr** can improve the performance of value decomposition MARL

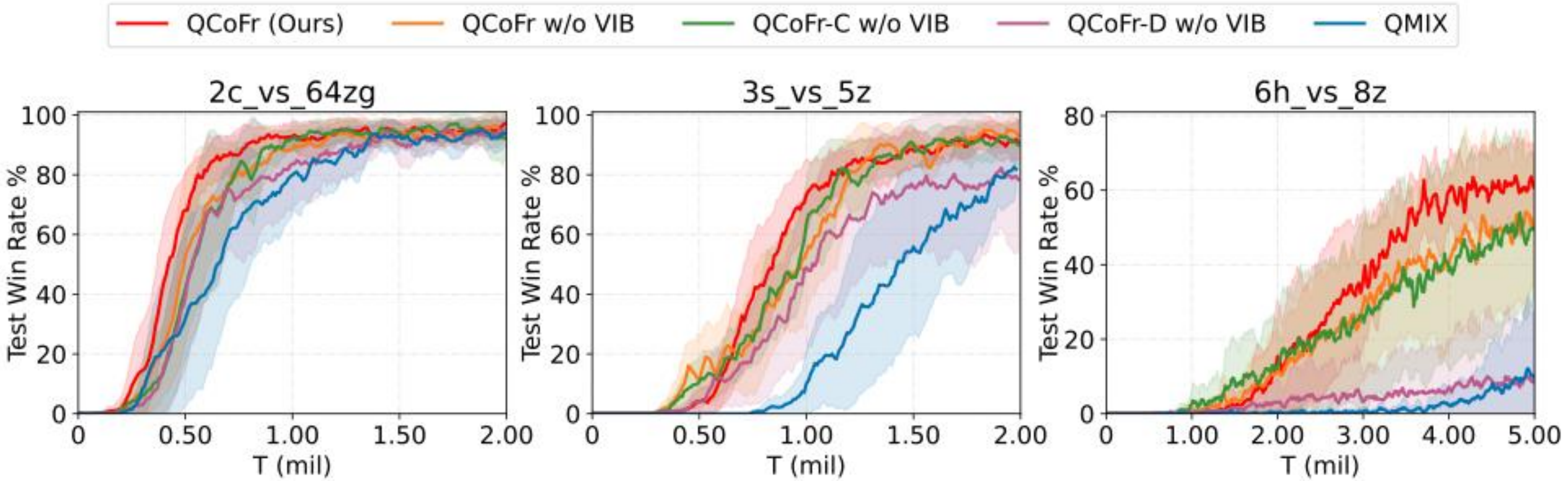


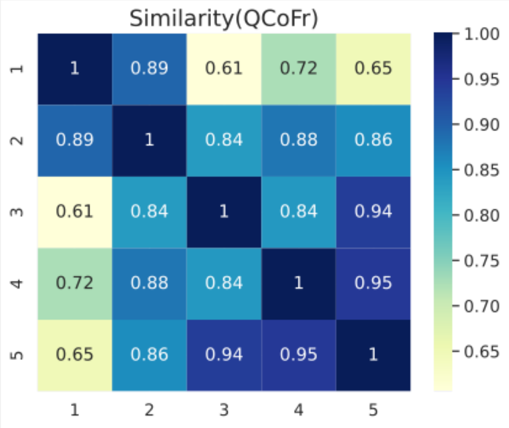
Experiments — Ablation Studies

- Modeling higher-order interactions improves coordination performance



- CFN efficiently captures high-order interactions, while VIB further enhances coordination





Interpretability

- **Meaningful coalitions**
  - ➢ Coordinated coalitions focusing fire
- **Higher-order cooperation**
  - ➢ Deeper CFN ⇒ More-agent collaborations
- **Role specialization**
  - ➢ Low-health agent (1) disengages
- **Diverse agent behaviors**
  - ➢ Lower cosine similarity ⇒ More specialized, diverse policies
- **More interpretable than black-box baselines**

Conclusion

- **QCoFr:** An interpretable, value-based MARL framework built on **CFN** and **VIB**
  - Explicitly models **arbitrary order agent interactions** with low complexity and clear attributions to individuals and coalitions
  - **Outperforms** strong value-decomposition baselines while enhancing **interpretability** and **coordination** analysis
- **Future work:**

  Develop **adaptive depth mechanisms** to dynamically adjust interaction order according to task complexity

Thanks for listening!