Interference Management Based on Deep Reinforcement Learning for MIMO Wireless Networks

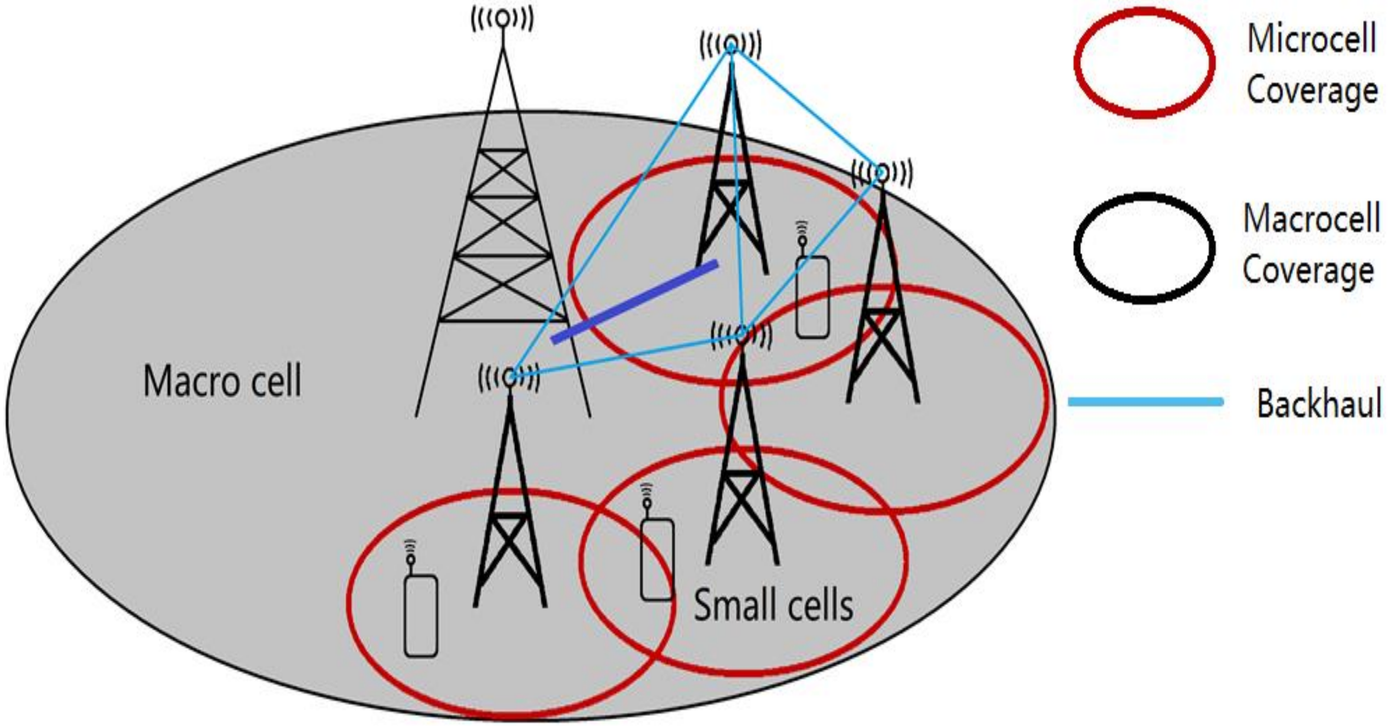
Fukawa Lab
LIU Yuchen 20M18917

Index

- Background
- Previous Researches
- System Model
- Key Ideas
- Conclusions
- References

Background: Interference Management (IM)

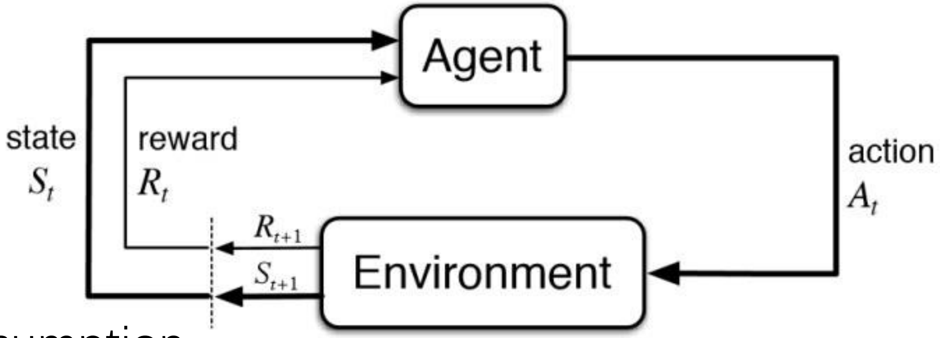
- Increasing of carrier frequency in mobile communication system
- Demand for high data traffic
- Demand of dense deployment of small cell wireless communication system
- **Interference Management (IM)** is essential to such systems



Advanced interference management for 5G cellular networks. Nam et, al. [1]

Background: Reinforcement Learning (RL)

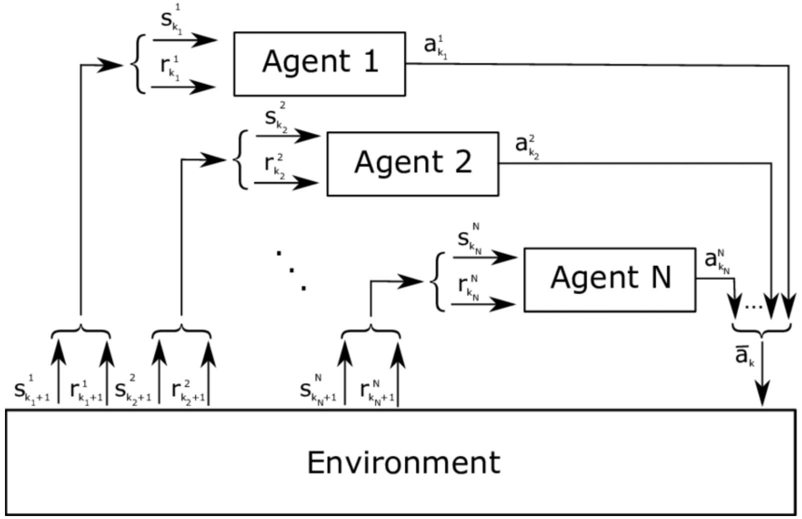
Single Agent Reinforcement Learning (SARL)
Deep reinforcement learning: An overview. Li [2]



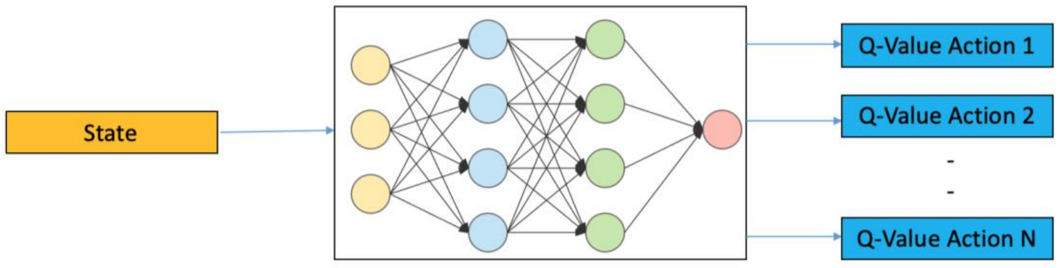
S_t: current environment state
A_t: agent action base on current state
R_t: Reward of action in current state
S_{t+1}, R_{t+1}: Next state and award, base on Marco model assumption

Multiple Agent Reinforcement Learning (MARL)
A comprehensive survey of multiagent reinforcement learning. Busoniu et al., [3]

There are multiple agents exist, and every agent interacts with environment and change environment.



Deep Q-Network

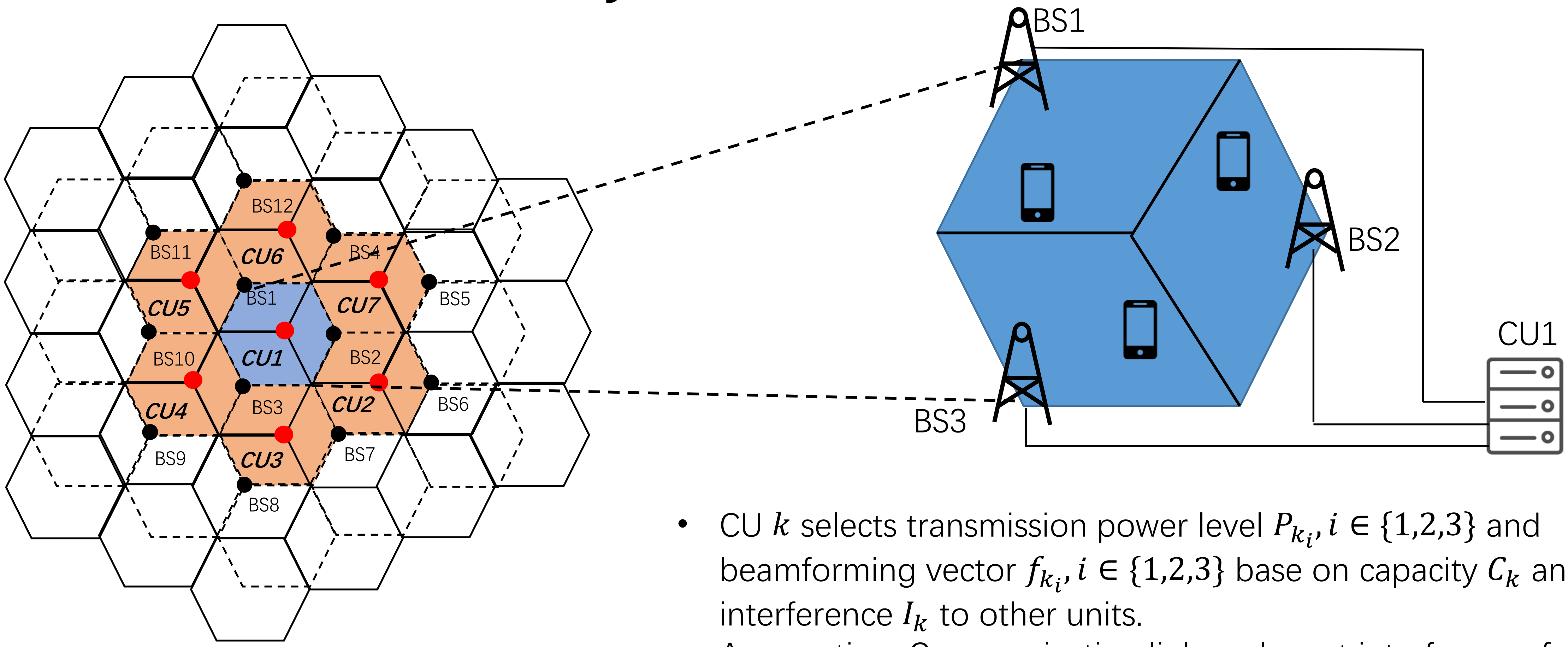


Deep Q Learning

Previous Researches

Research	Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks. Nasir et al., [4]	Distributed Inter-cell Interference Coordination for Small Cell Wireless Communications: A Multi-Agent Deep Q-Learning Approach. Jiang et al., [5]	Deep Reinforcement Learning for Distributed Dynamic MISO Downlink-Beamforming Coordination. Liang et al., [6]
Control Elements	Power Level	Power Level, Beamforming Vector	Power Level, Beamforming Vector
Train & Execute	Centralized Training Distributed Execute	Centralized Training Distributed Execute	Distributed Training Distributed Execute
Q-Network	Single Network	Multiple Network	Multiple Network
communication	Yes	No	Yes
Baseline	Fractional Programming, Full Power, Random	Exhaustive Search, Random, Neural Network	Fractional Programming, Greedy, Random
DQN Train	Time Difference	Monte-Carlo	Time Difference
Deficient	Beamforming is not considered; Training is centralized manner	Each agent need all channel state information	Intuitive design state input and exchanged information

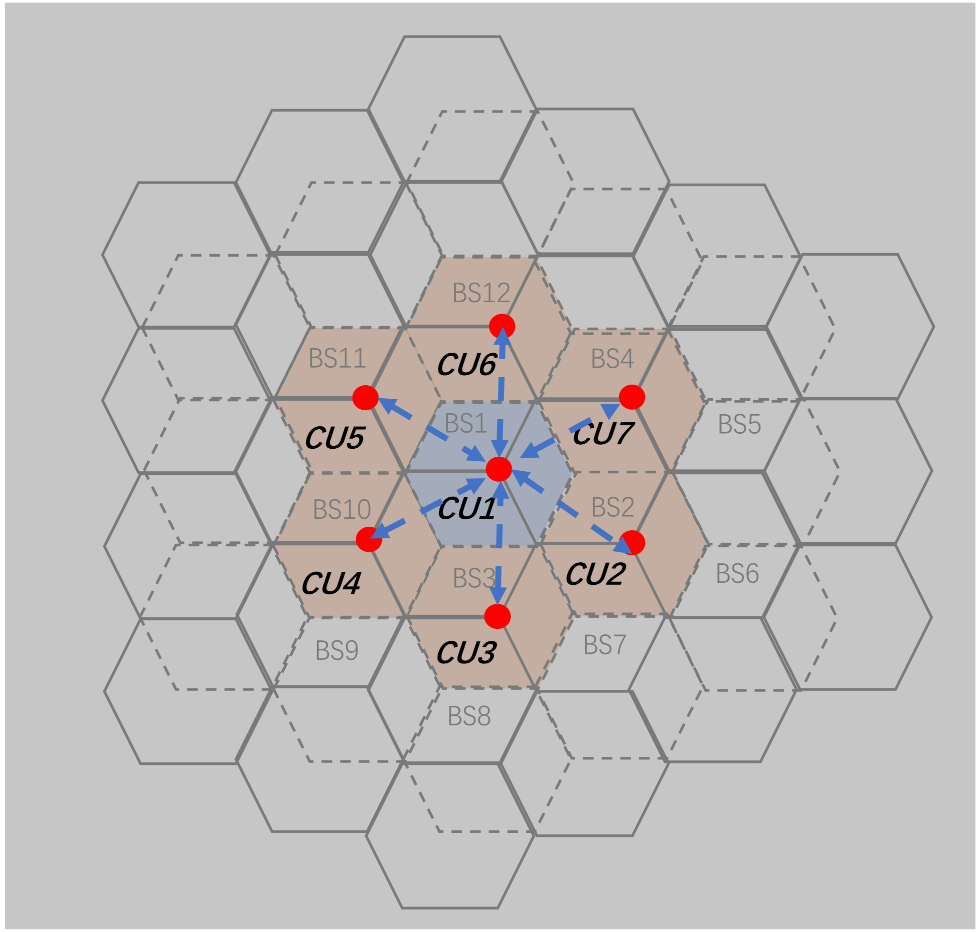
System Model: Multi-Sectors Small-Cell Wireless Communication System



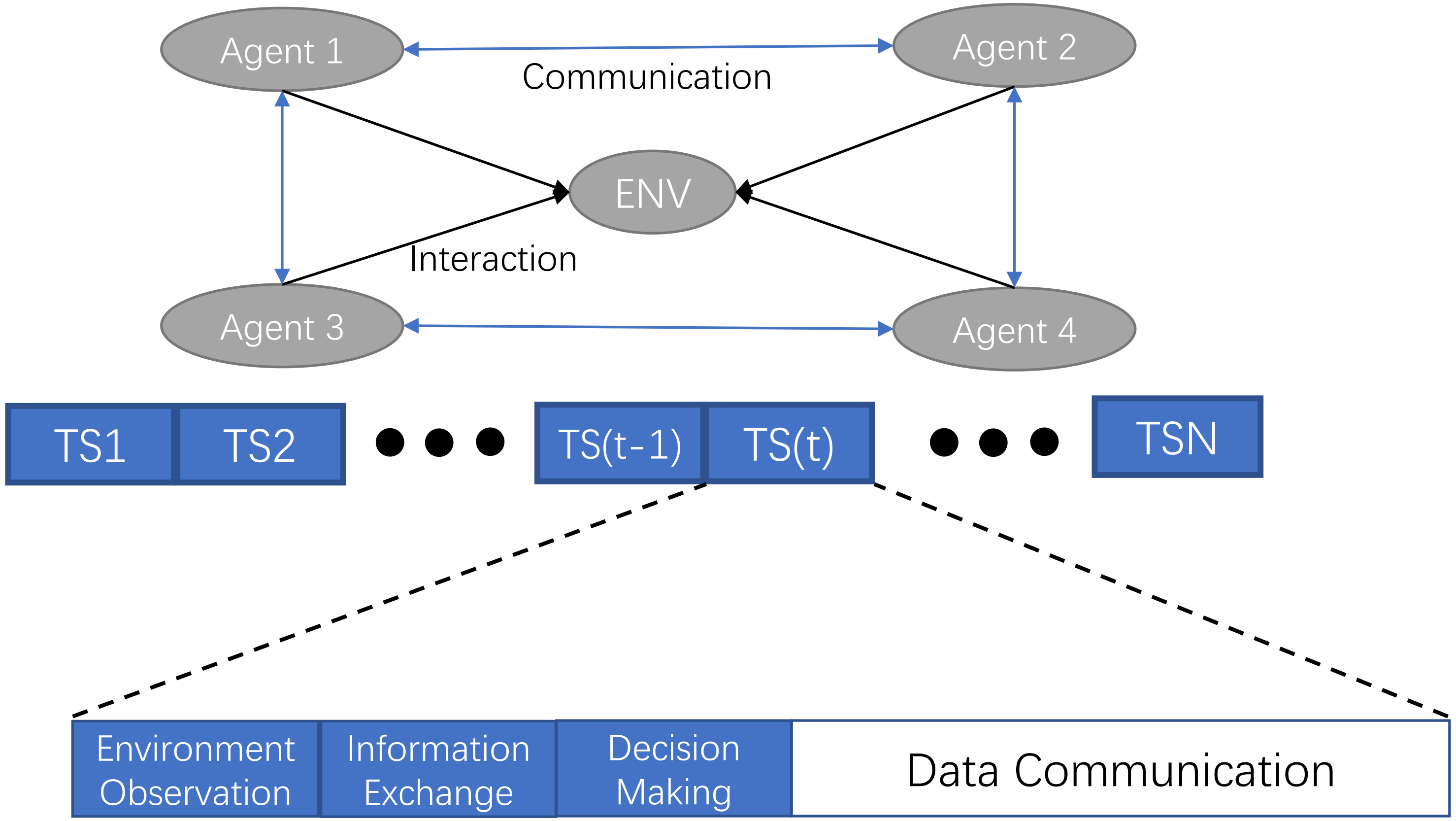
- Coordination Unit (CU)
- Base Station (BS)

- CU k selects transmission power level $P_{k_i}, i \in \{1,2,3\}$ and beamforming vector $f_{k_i}, i \in \{1,2,3\}$ base on capacity C_k and interference I_k to other units.
- Assumption: Communication links only get interference from adjoin neighboring CUs.

Key Ideas: Information Exchange



MARL has problem of non- stationarity in environment, information exchange is one of major way to keep the stationery of environment.



Key Ideas: Channel Capacity and Optimization Problem

Capacity of CU k

Interference from sectors of current CU

Interference from other CUs (Not able to get this directly)

$$C_k = \frac{1}{3}\sum_{i=1}^{3} C_{k_i}$$

$$= \frac{1}{3}\sum_{i=1}^{3} \log_2\left(1 + \frac{p_{k_i}||\boldsymbol{H}_{i,i}\boldsymbol{f}_i||^4}{\sigma_n^2||\boldsymbol{H}_{i,i}\boldsymbol{f}_i||^2 + \boxed{\sum_{j=1,j\neq i}^{3} p_j||\boldsymbol{f}_j^H\boldsymbol{H}_{j,i}^H\boldsymbol{H}_{j,i}\boldsymbol{f}_j||^2} + \boxed{\sum_{l=1}^{6}\sum_{n=1}^{3} p_n^l {\boldsymbol{f}_n^l}^H {\boldsymbol{H}_{n,i}^l}^H \boldsymbol{H}_{n,i}^l \boldsymbol{f}_n^l}}\right)$$

Optimization Problem

$$\mathcal{P}: \quad \text{argmax}: C = \frac{1}{K}\sum_{k=1}^{K} C_k$$

$$s.t.\ p_n^l \in \{p_0, p_1, \ldots, p_{Q_{pow}-1}\}$$

$$\boldsymbol{f}_n^l \in \{\boldsymbol{f}_0, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_{Q_{beam}-1}\}$$

$\boldsymbol{f}_n^l$: Beamforming vector of n-th sector in l-th CU

p_n^l: Power level of of n-th sector in l-th CU

$H_{i,j}^l$: Channel Matrix from i-th sector of current CU to j-th sector of l CU

α: Interference penalty parameter

Revised CU Capacity to Other CUs

Interference Penalty

$$C_k' = \sum_{i}^{3} C_{k_i}\left(1 - \boxed{\alpha\sum_{l=1}^{6}\sum_{n=1}^{3} p_n^l \boldsymbol{f}_i^H {\boldsymbol{H}_{i,n}^l}^H \boldsymbol{H}_{i,n}^l \boldsymbol{f}_i}\right)$$

Key Ideas: State, Action and Reward

Action Space of of CU k

$$Q_{CU} = Q_{pow} \times Q_{beam} \times 3$$

$$\mathcal{A} = \{(p, f), p \in \mathcal{P}, f \in \mathcal{F}\}$$

$$\mathcal{P} = \{0, \frac{1}{Q_{pow} - 1} p_{max}, \frac{2}{Q_{pow} - 1} p_{max}, \dots, p_{max}\}$$

$$\mathcal{F} = \{\boldsymbol{f}_0, \boldsymbol{f}_1, \dots, \boldsymbol{f}_{Q_{beam}-1}\}$$

State Space of of CU k

$$S_k(t) = \begin{cases} \text{Loacal Information:} & H_{i,j}^k(t), H_{j,i}^k(t) \\ \text{Exchange Information:} & \begin{cases} H_{j,i}^l(t), H_{i,j}^l(t), l \in \{1,2,\dots,6\} \\ p_n^l(t-1), \boldsymbol{f}_n^l(t-1), \\ l \in \{1,2,..,6\}, n \in \{1,2,3\} \end{cases} \end{cases}$$

Reward Function of of CU k

Interference from Same CU

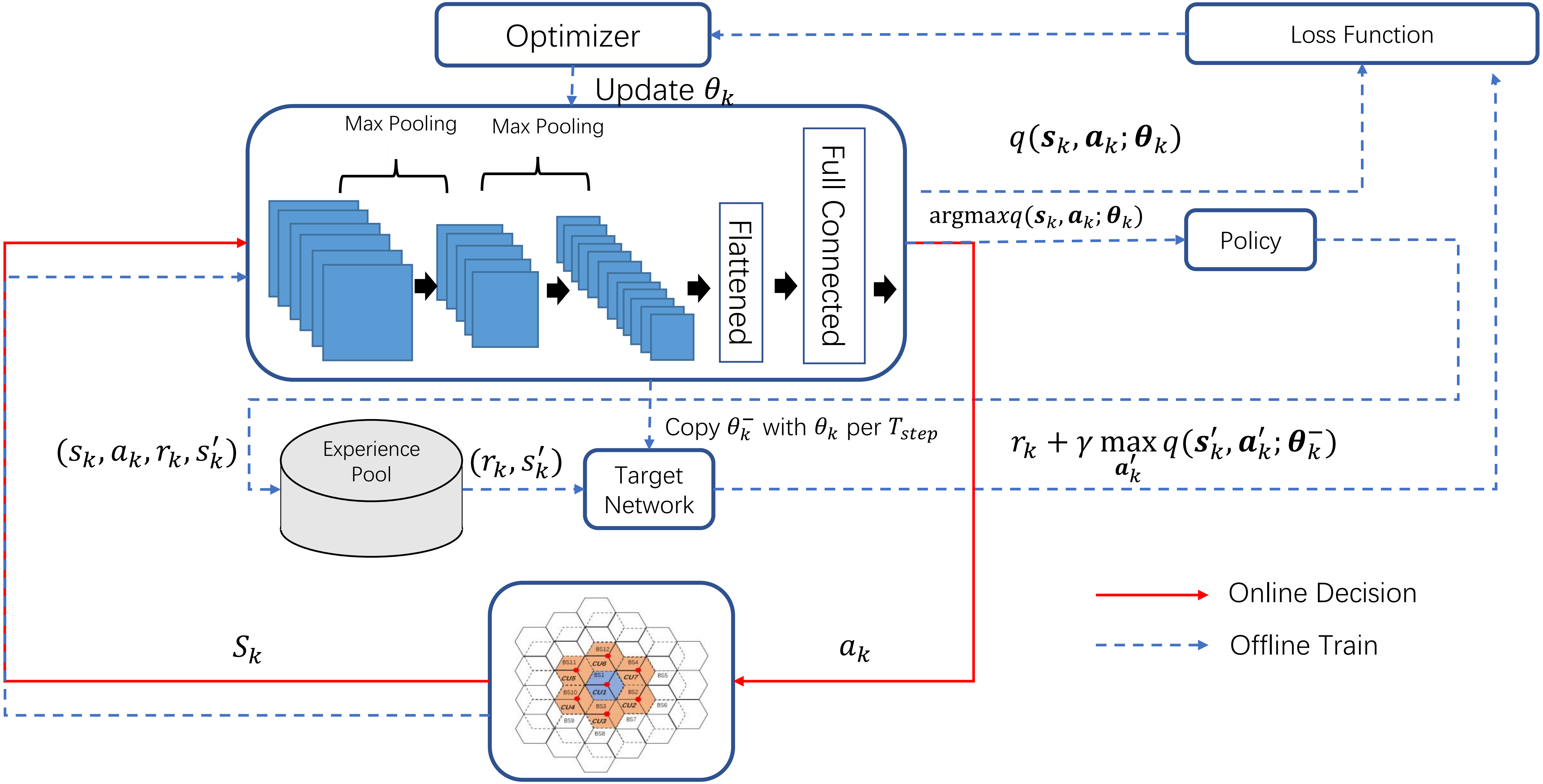
Interference from Other CUs

$$R_k(t) = \sum_{i}^{3} \log_2\left(1 + \frac{p_{k_i}(t)||\boldsymbol{H}_{i,i}(t)\boldsymbol{f}_i(t)||^4}{\sigma_n^2||\boldsymbol{H}_{i,i}(t)\boldsymbol{f}_i(t)||^2 + \boxed{\sum_{j=1, j\neq i}^{3} p_j(t)||\boldsymbol{f}_j^H(t)\boldsymbol{H}_{j,i}^H(t)\boldsymbol{H}_{j,i}(t)\boldsymbol{f}_j(t)||^2} + \boxed{\sum_{l=1}^{6}\sum_{n=1}^{3} p_n^l(t-1)\boldsymbol{f}_n^l(t-1)^H\boldsymbol{H}_{n,i}^l(t)^H\boldsymbol{H}_{n,i}^l(t)\boldsymbol{f}_n^l(t-1)}}\right)$$

$$\times \boxed{(1 - \alpha \sum_{l=1}^{6}\sum_{n=1}^{3} p_n^l(t)\boldsymbol{f}_i^H(t)\boldsymbol{H}_{i,n}^l(t)^H\boldsymbol{H}_{i,n}^l(t)\boldsymbol{f}_i(t))}$$

Interference Penalty

Key Ideas: Convolution Q-Network



Conclusions

- Designing a system model of 3-sectors small-cell cellular system
- Interference management combines **centralized and distributed manners** (centralized manner in CU, distributed manner among agents)
- Adopting **information exchange** among agents to hold the stationarity of environment for MARL algorithm convergence
- Proposing **convolution Q-network** for channel state information input
- Reward function contains **interference penalty** to other CUs

References

[1] Nam, W., Bai, D., Lee, J., & Kang, I. (2014). Advanced interference management for 5G cellular networks. IEEE Communications Magazine, 52(5), 52-60.

[2] Li, Y. (2017). Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274.

[3] Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 38(2), 156-172.

[4] Nasir, Y. S., & Guo, D. (2019). Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks. IEEE Journal on Selected Areas in Communications, 37(10), 2239-2250.

[5] Jiang, S., Chang, Y., & Fukawa, K. (2020, October). Distributed Inter-cell Interference Coordination for Small Cell Wireless Communications: A Multi-Agent Deep Q-Learning Approach. In *2020 International Conference on Computer, Information and Telecommunication Systems (CITS)* (pp. 1-5). IEEE.

[6] Ge, J., Liang, Y. C., Joung, J., & Sun, S. (2020). Deep Reinforcement Learning for Distributed Dynamic MISO Downlink-Beamforming Coordination. *IEEE Transactions on Communications*, *68*(10), 6070-6085.

[7] Dong, J., Chen, S., Ha, P. Y. J., Li, Y., & Labi, S. (2020). A DRL-based Multiagent Cooperative Control Framework for CAV Networks: a Graphic Convolution Q Network. arXiv preprint arXiv:2010.05437.

Thank You for Listening
ご清聴ありがとうございました

Q&A

Appendix

Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks[1]

$$\begin{aligned} &\underset{\boldsymbol{p}}{\text{maximize}} \quad \sum_{i=1}^{n} w_i^{(t)} \cdot C_i^{(t)}(\boldsymbol{p}) \\ &\text{subject to} \quad 0 \le p_i \le P_{\max}, \quad i = 1, \ldots, n, \end{aligned} \tag{5}$$

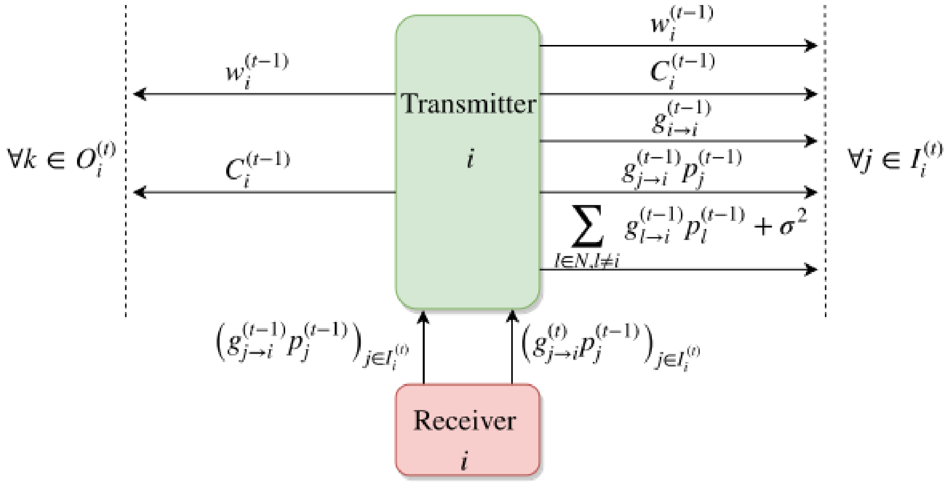


Fig. 1. The information exchange between transmitter i and its neighbors in time slot $t-1$. Note that transmitter i obtains $g_{j\to i}^{(t)} p_j^{(t-1)}$ by the end of slot $t-1$, but it is not able to deliver this information to interferer j before the beginning of slot t due to additional delays through the backhaul network.

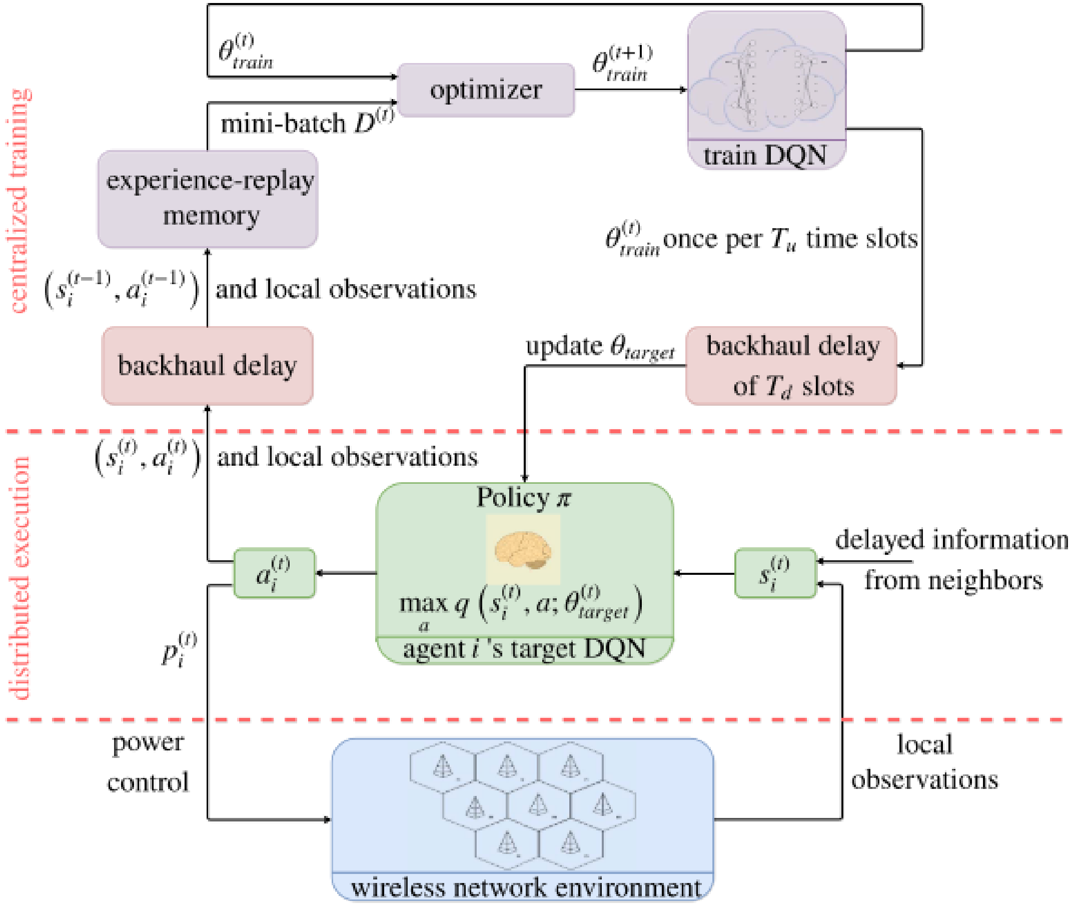


Fig. 2. Illustration of the proposed multi-agent deep reinforcement learning algorithm.

Information exchange among multi-agent to conduct a centralized training distributed (CTDE) execute power allocation method.

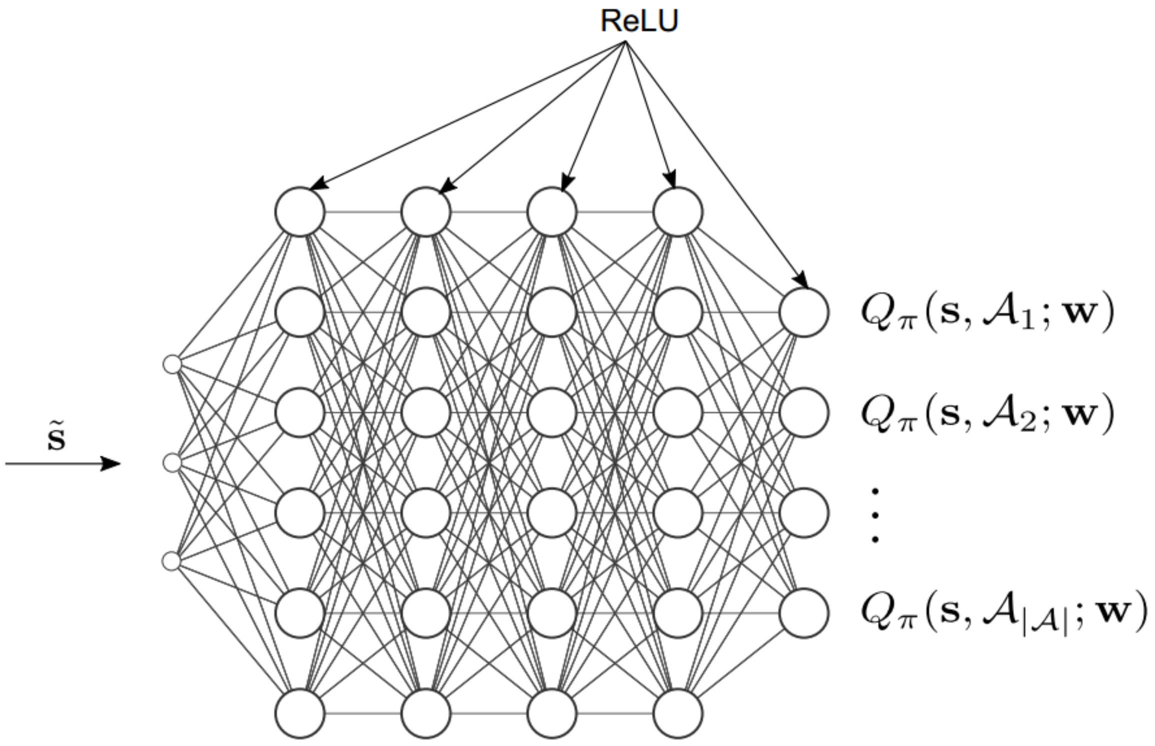
Distributed Inter-cell Interference Coordination for Small Cell Wireless Communications: A Multi-Agent Deep Q-Learning Approach[2]

$$C = \frac{1}{K} \sum_{i=1}^{K} \log_2(1 + \mathrm{SINR}_i). \tag{7}$$

$$\begin{aligned} \mathcal{P}: \quad & \max \ C \\ s.t. \ & \mathbf{f}_k \in \mathbb{F}, \forall k \in [1, K], \text{and} \\ & p_k \in \mathbb{P}, \forall k \in [1, K]. \end{aligned} \tag{8}$$

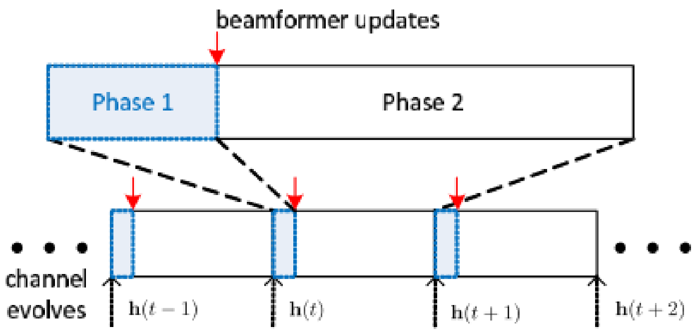
$$\mathbf{x}_{i,j} = \left[||\mathbf{g}_{i,j}(\mathbb{F}_1)||^2, \ldots, ||\mathbf{g}_{i,j}(\mathbb{F}_{|\mathbb{F}|})||^2 \right]^{\mathrm{T}}, \tag{15a}$$

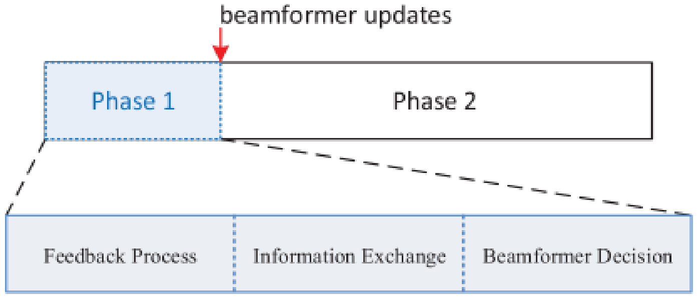
$$\mathbf{s} = \left[\mathbf{x}_{1,1}^{\mathrm{T}}, \ldots, \mathbf{x}_{1,K}^{\mathrm{T}}, \ldots, \mathbf{x}_{K,K}^{\mathrm{T}} \right]^{\mathrm{T}}, \tag{15b}$$

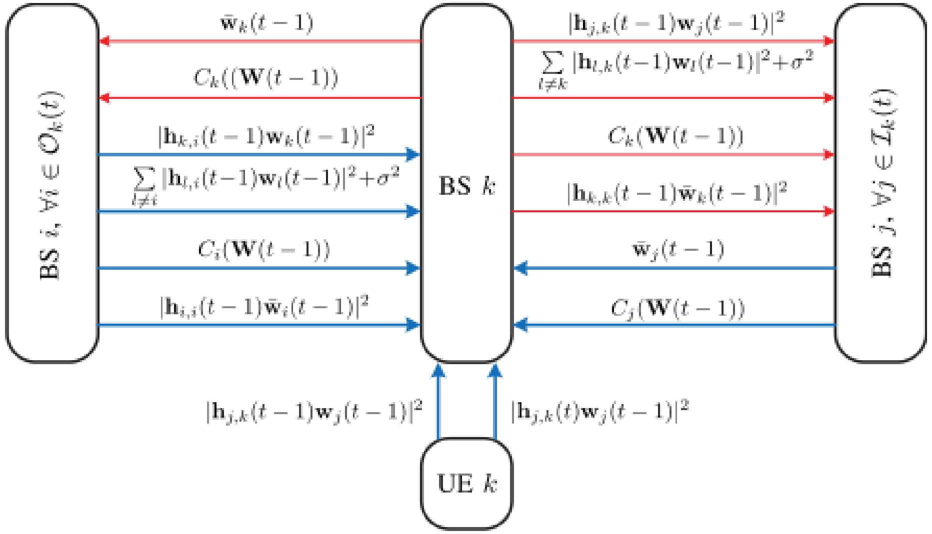


Joint Optimization of power level and beamforming vector selection in wireless network. Each agent train independent Q-network for distributed execution. However, overall channel state information (CSI) is needed for state input of Q-network.

Deep Reinforcement Learning for Distributed Dynamic MISO Downlink-Beamforming Coordination[3]

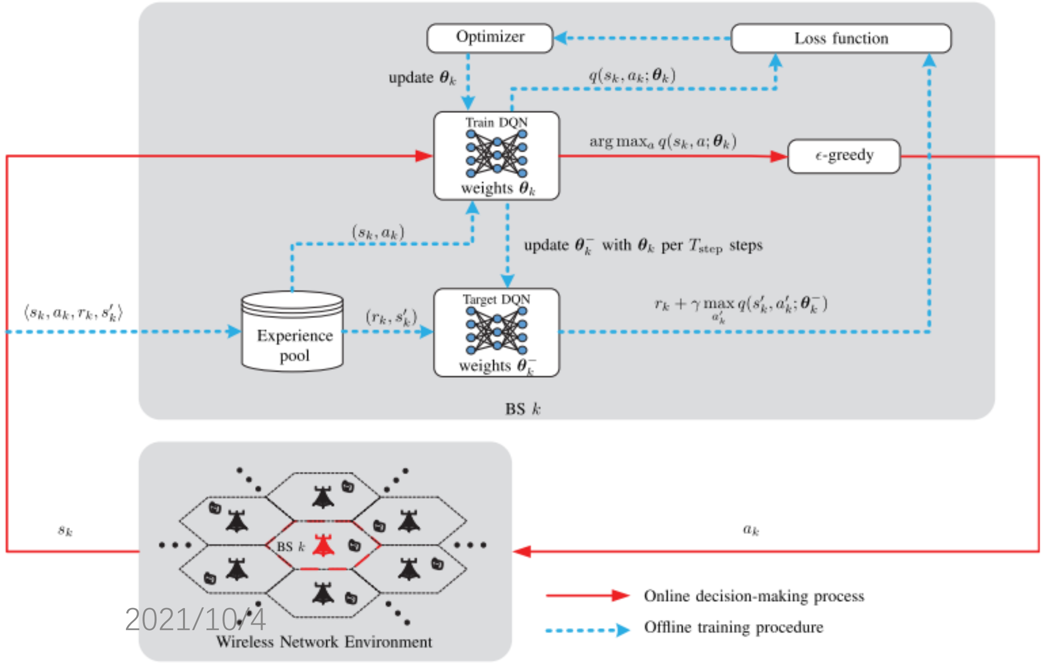






$$\max_{\mathbf{W}(t)} \sum_{k=1}^{K} C_k(\mathbf{W}(t))$$

$$\text{s.t. } 0 \le \|\mathbf{w}_k(t)\|^2 \le p_{\max}, \ \forall k \in \mathcal{K}, \qquad (8b)$$



This research design an information exchange protocol to exchange information in last time slot like beamforming decision, single cell capacity and interference, etc. Instead of using full CSI as state input, they design an intuitive state input base on local and exchange information. Finally, a distributed train distributed execute algorithm is proposed.