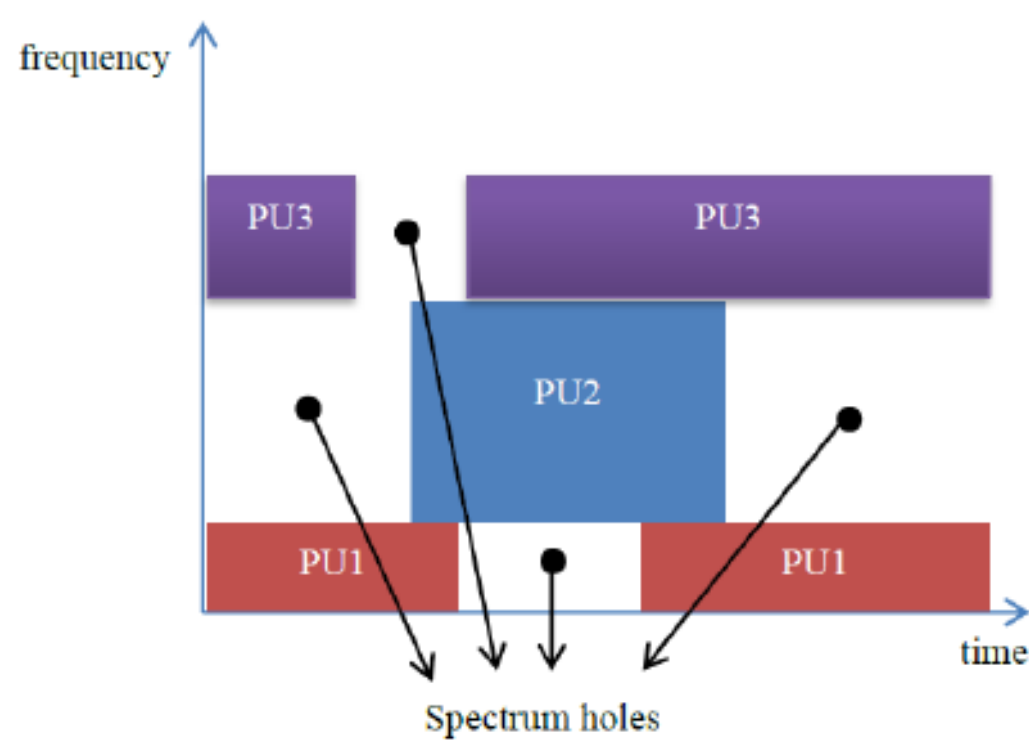


Deep Reinforcement Learning For Dynamic Multichannel Access

Shangxing Wang and Bhaskar Krishnamachari, EE/ANRG

Motivation I

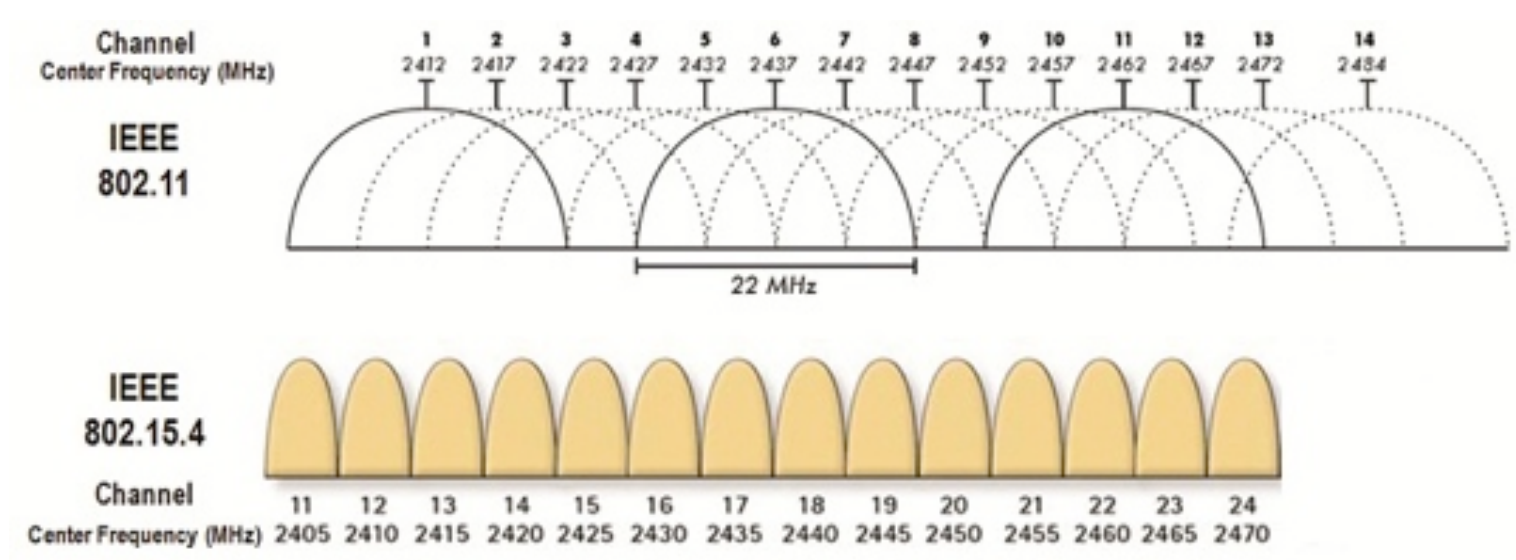
- Wireless network has an exponentially increasing data traffic, but limited spectrum
- Fixed spectrum assignment causes spectrum inefficiency



- Dynamic spectrum access** allows SUs to use idle channels to improve spectrum utilization

Motivation II

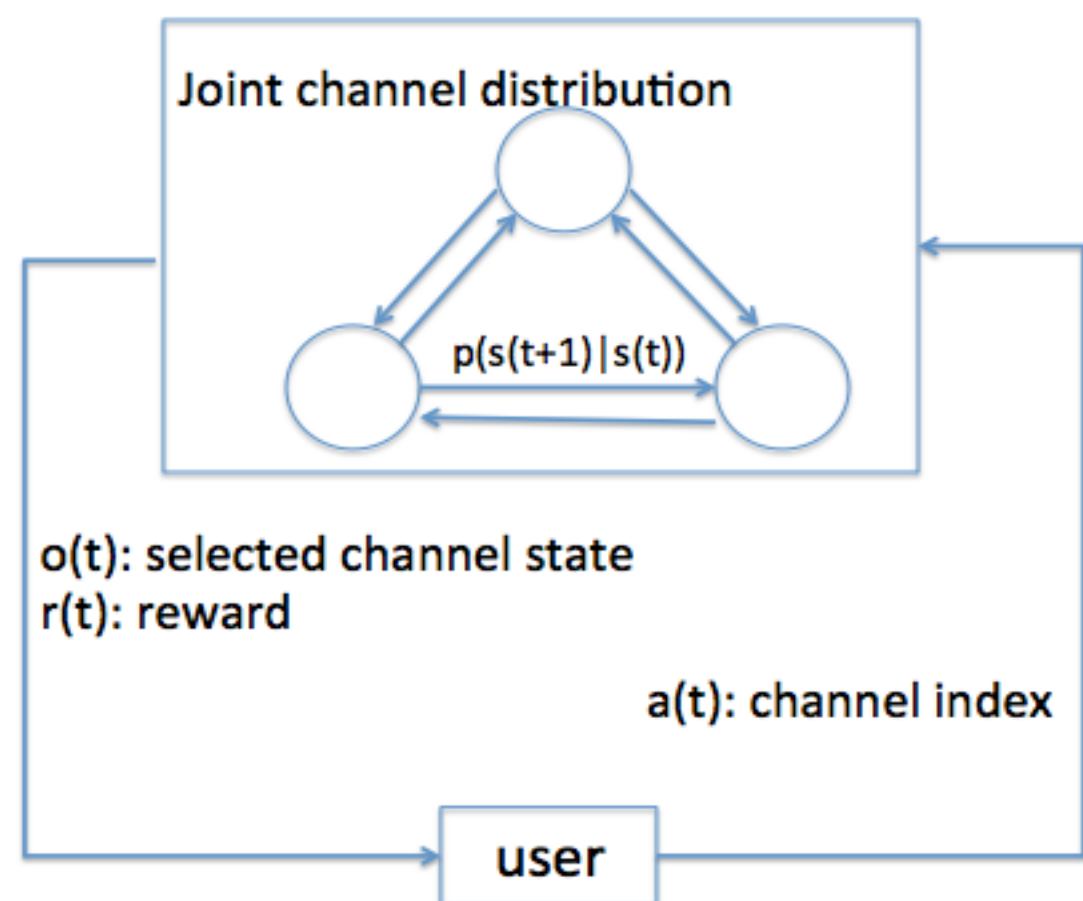
- Highly correlated channels in real systems



- Goal: design a dynamic multichannel access policy for **correlated** systems with optimal/near-optimal performance

Problem Formulation

- POMDP



- Objective: design a channel access policy to maximize the long term expected accumulated discounted reward
- Challenge:** computationally intractable unknown system dynamics

Approach

- Q learning: an **online** approach that learns the policy directly

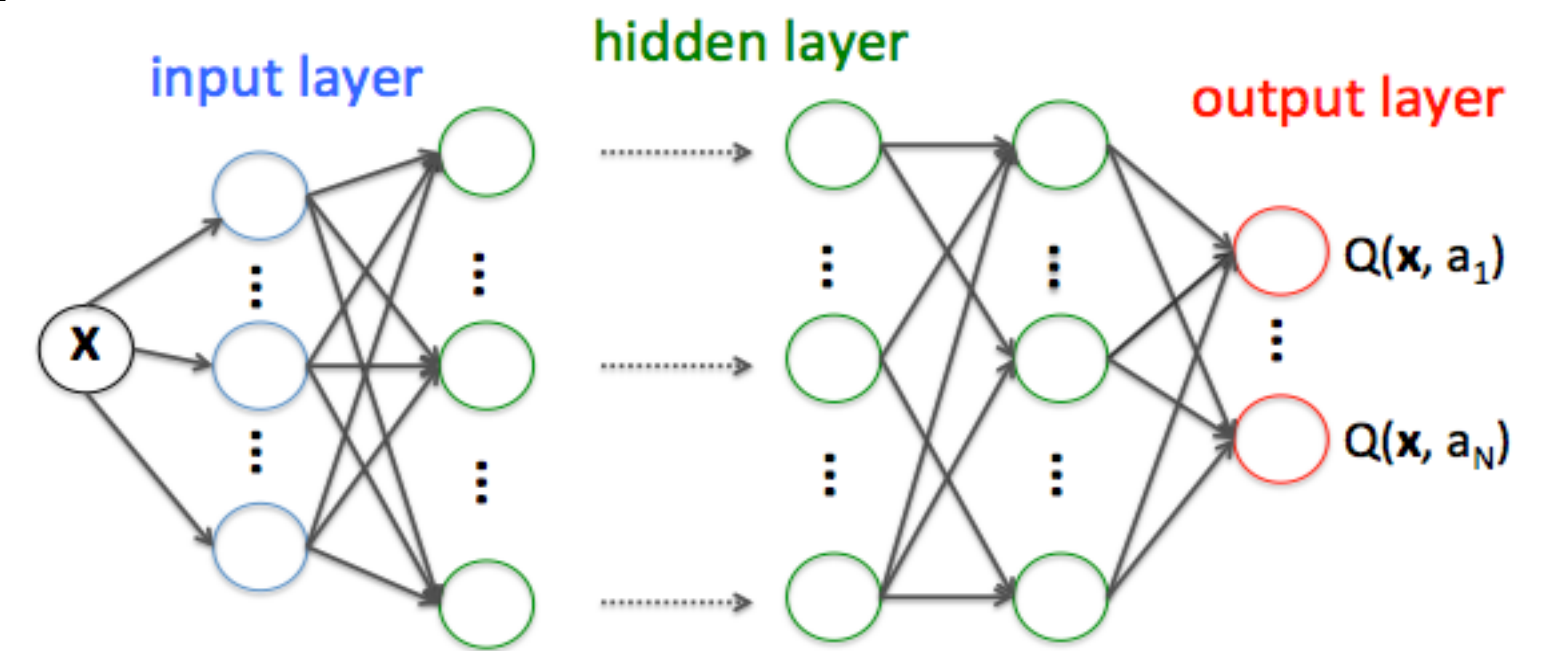
- State: $\mathbf{x}_t = [a_{t-1}, o_{t-1}, \dots, a_{t-M}, o_{t-M}]$.

- Q value update rule:

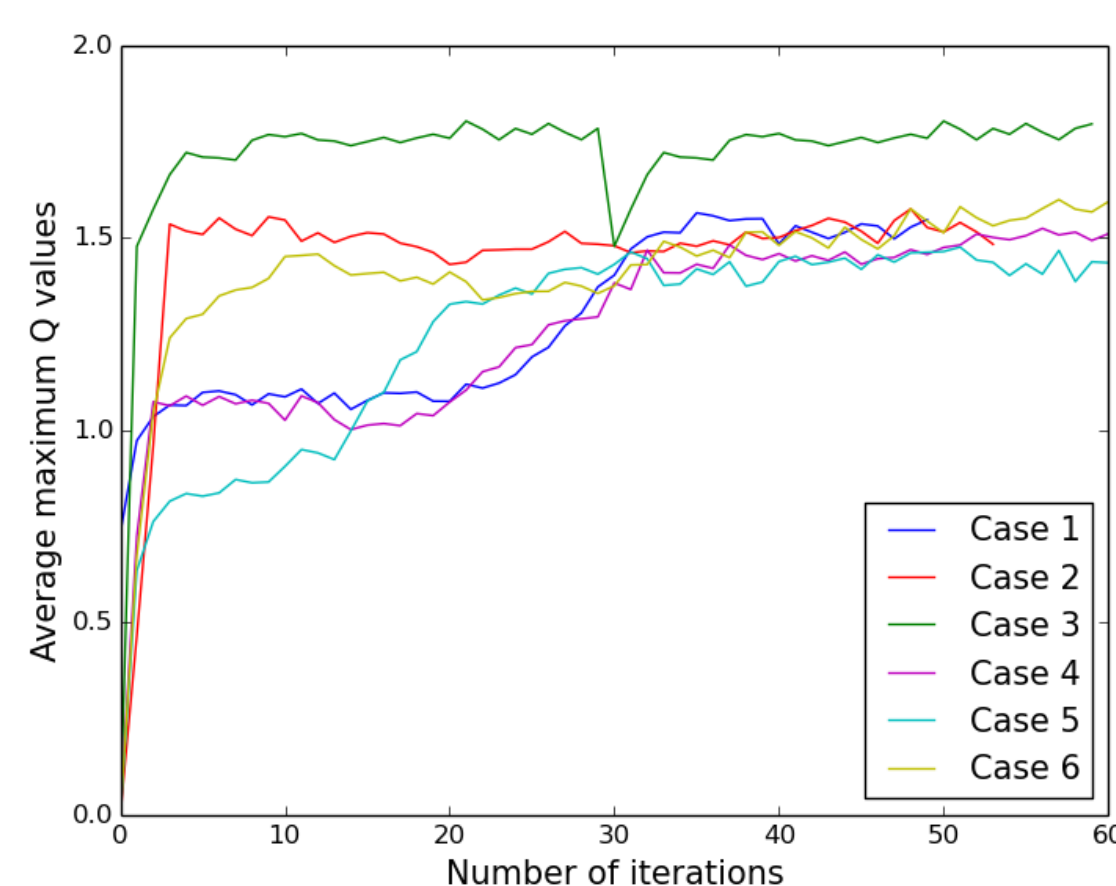
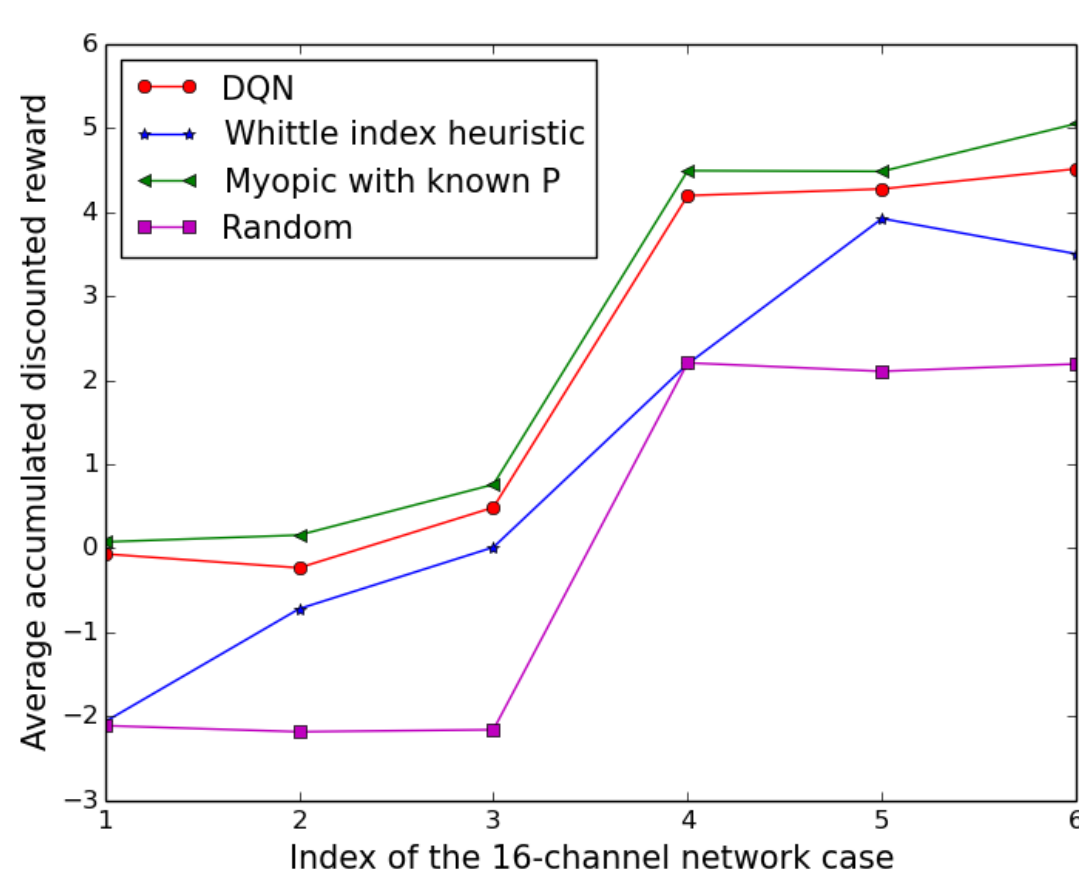
$$Q(\mathbf{x}_t, a_t) \leftarrow Q(\mathbf{x}_t, a_t) + \alpha [r_{t+1} + \gamma \max_{a_{t+1}} Q(\mathbf{x}_{t+1}, a_{t+1}) - Q(\mathbf{x}_t, a_t)]$$

- Policy: $\pi^*(\mathbf{x}) = \arg \max_a Q^{\pi^*}(\mathbf{x}, a), \forall \mathbf{x}$.

- Deep Q Network:**

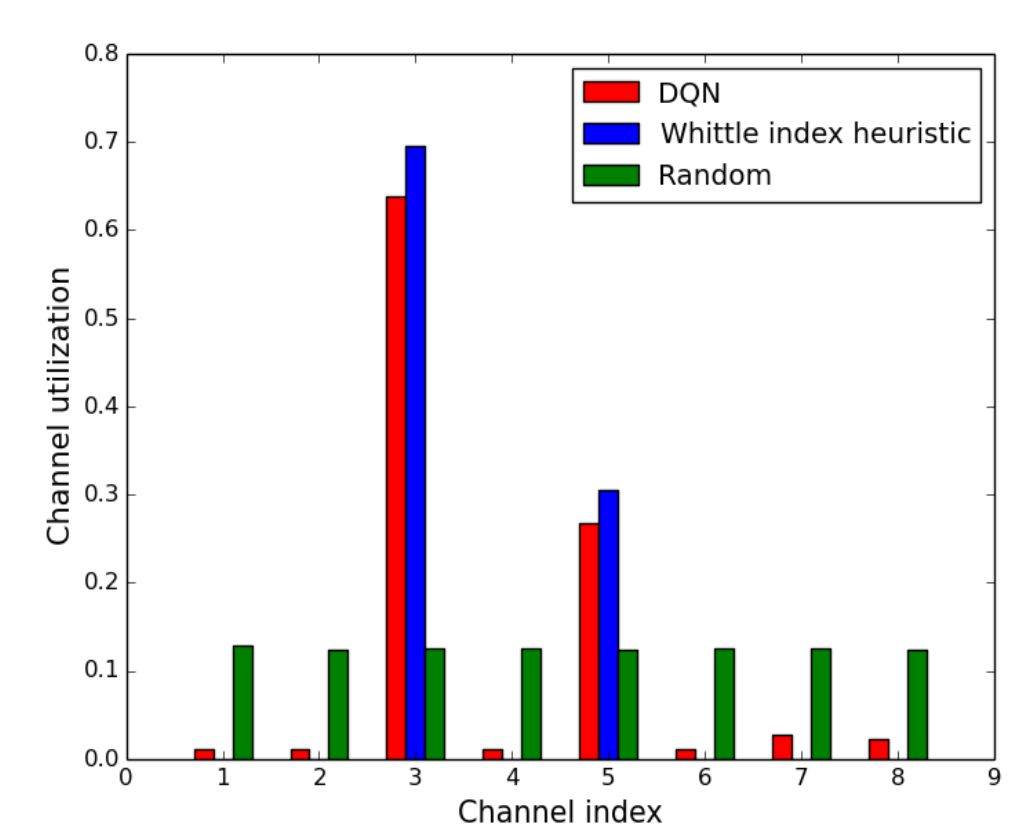


Evaluation



Simulation

policy	performance
DQN	0.947 (1.44)
Whittle Index	0.767 (1.35)
Random Policy	-2.170 (1)



Real Data Trace