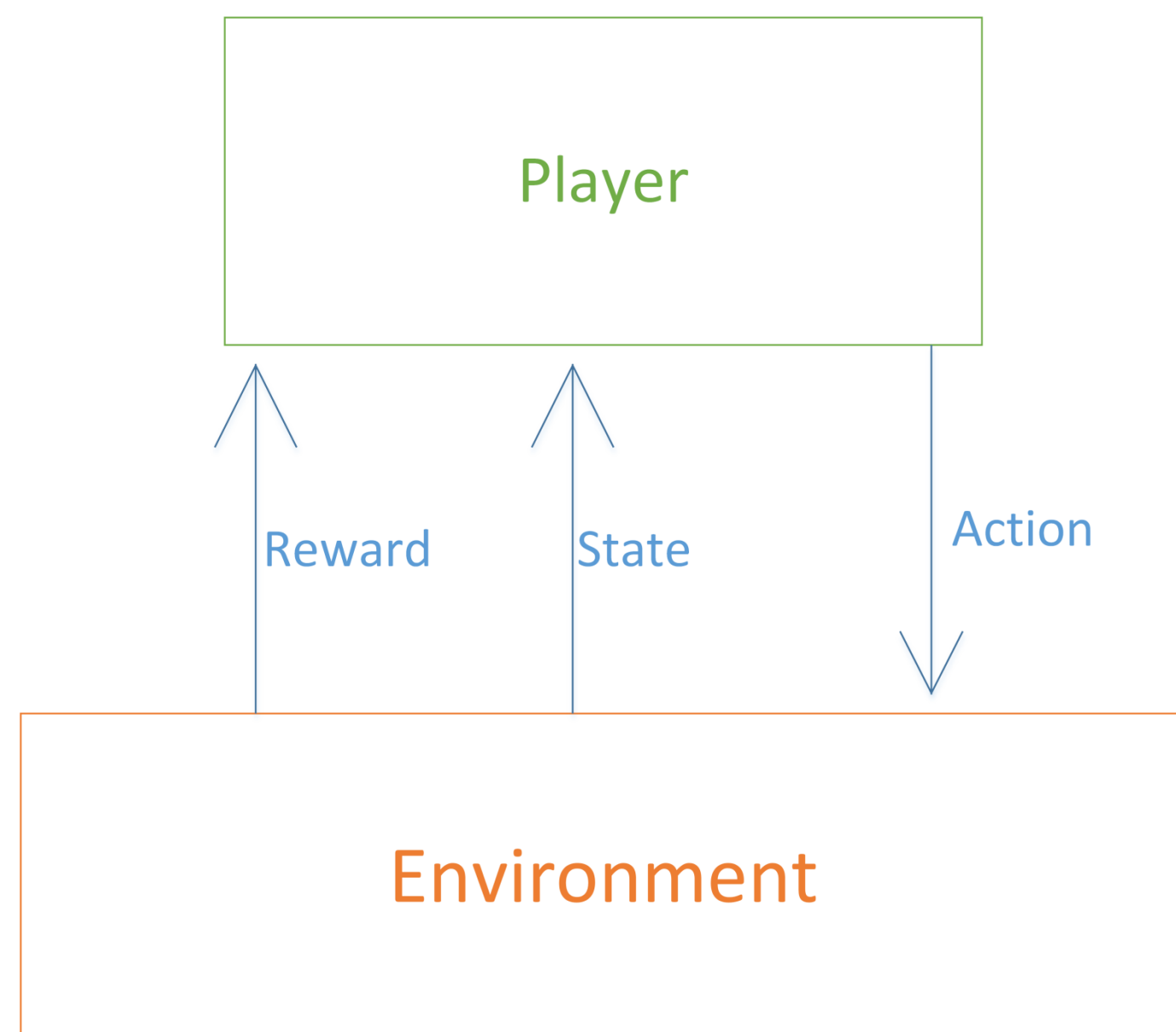


Adaptive Power and Rate Allocation over Markovian Fading Channels

Nachikethas A. Jagadeesan, Bhaskar Krishnamachari

Reinforcement Learning



- Map situations to rewards signals to maximize a reward signal
- Discover actions that yield the most reward by trying them
- A *policy* defines the learning player's behavior at any time slot
- A *reward function* defines the goal in a reinforcement learning (RL) problem
- The player's sole objective is to maximize the total reward in the long run
- How well a RL system works depends on how well it can generalize from past experience

Q-Learning

- Q-Learning (QL) is an example of a *temporal difference* (TD) learning method
- A simple TD method for determining the *value* of a state s is given by:

$$V(s) \leftarrow V(s) + \alpha [V(s') - V(s)]$$

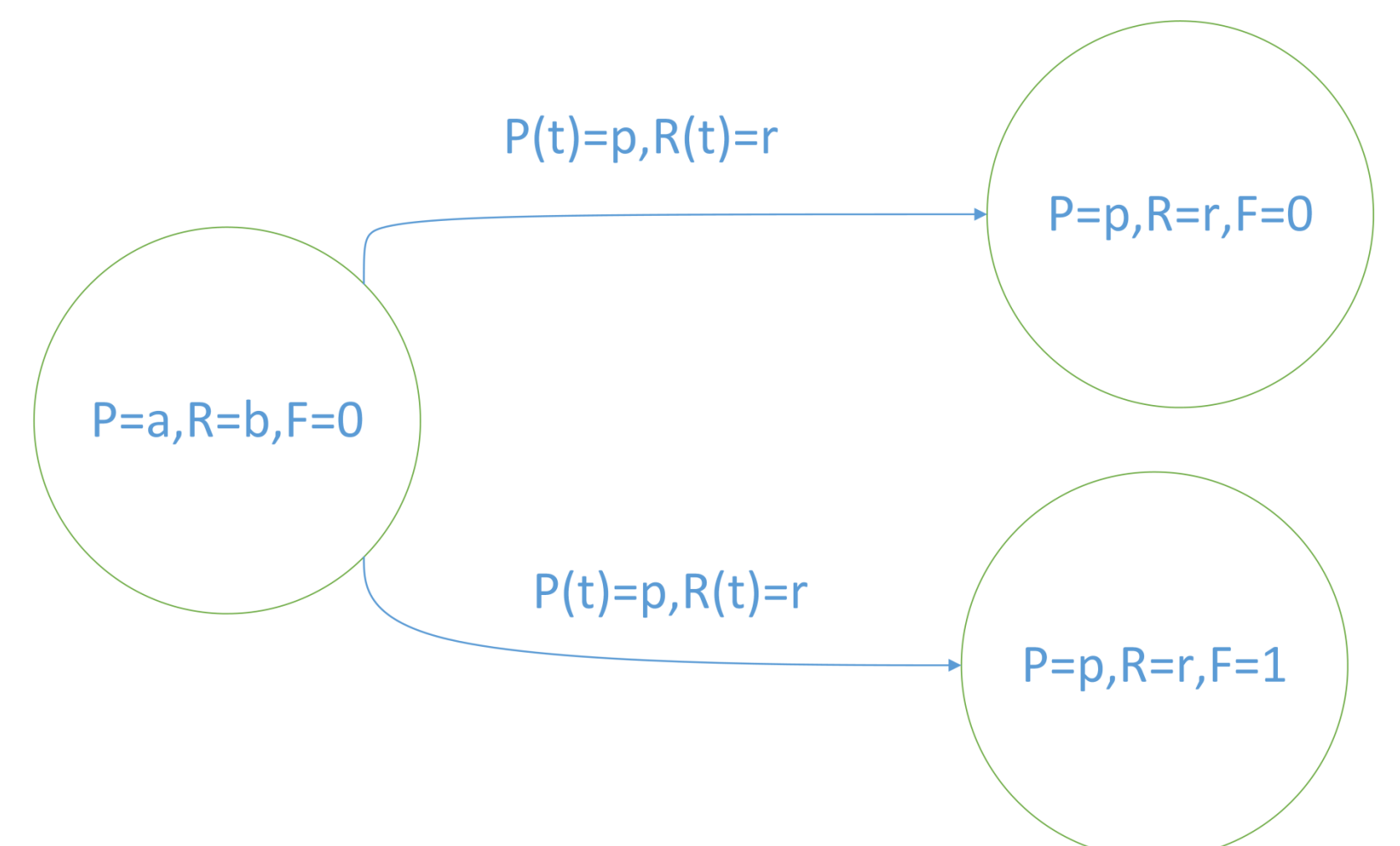
where α is a parameter that determines the rate of learning

- Consider a player who tries to maximize the expected discounted return given by $\sum_{k=0}^{\infty} \gamma^k R_{k+1}$
- QL maintains a function, Q , to approximate the optimal action-value function.

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

Markov Model

- Each state is defined by the 3-tuple (P, R, F) consisting of power, rate and feedback
- An action at time t , $a_{ij}(t)$ consists of a power-rate allocation of the form $(P(t), R(t))$



- The resulting state is then given by $(P(t), R(t), F(t))$ where $F(t)$ is the feedback from the channel for the transmission at time t
- There is a cost associated with each state action pair. In this model, the cost function is given by $C = R \times F - \beta P$
- Simulate the Q-Learning algorithm over a Nakagami channel model to evaluate performance.
- Implementation of the optimal policy on the WARP radio test bed is ongoing.

