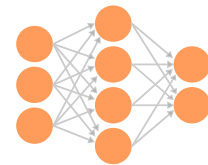


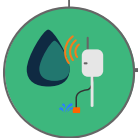
Agent

State (s_t)
Reward (r_t)

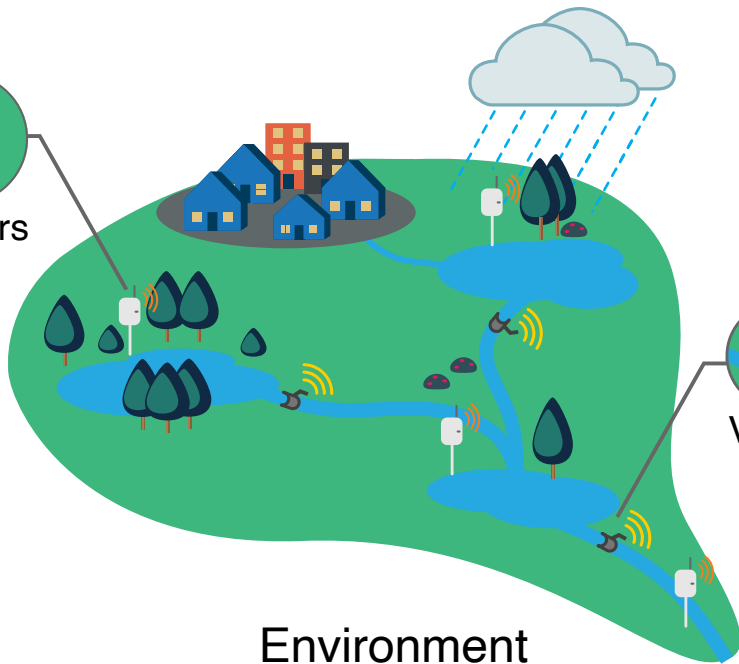


Policy $\pi(a_t|s_t)$

Actions (a_t)



Sensors



Environment



Valves