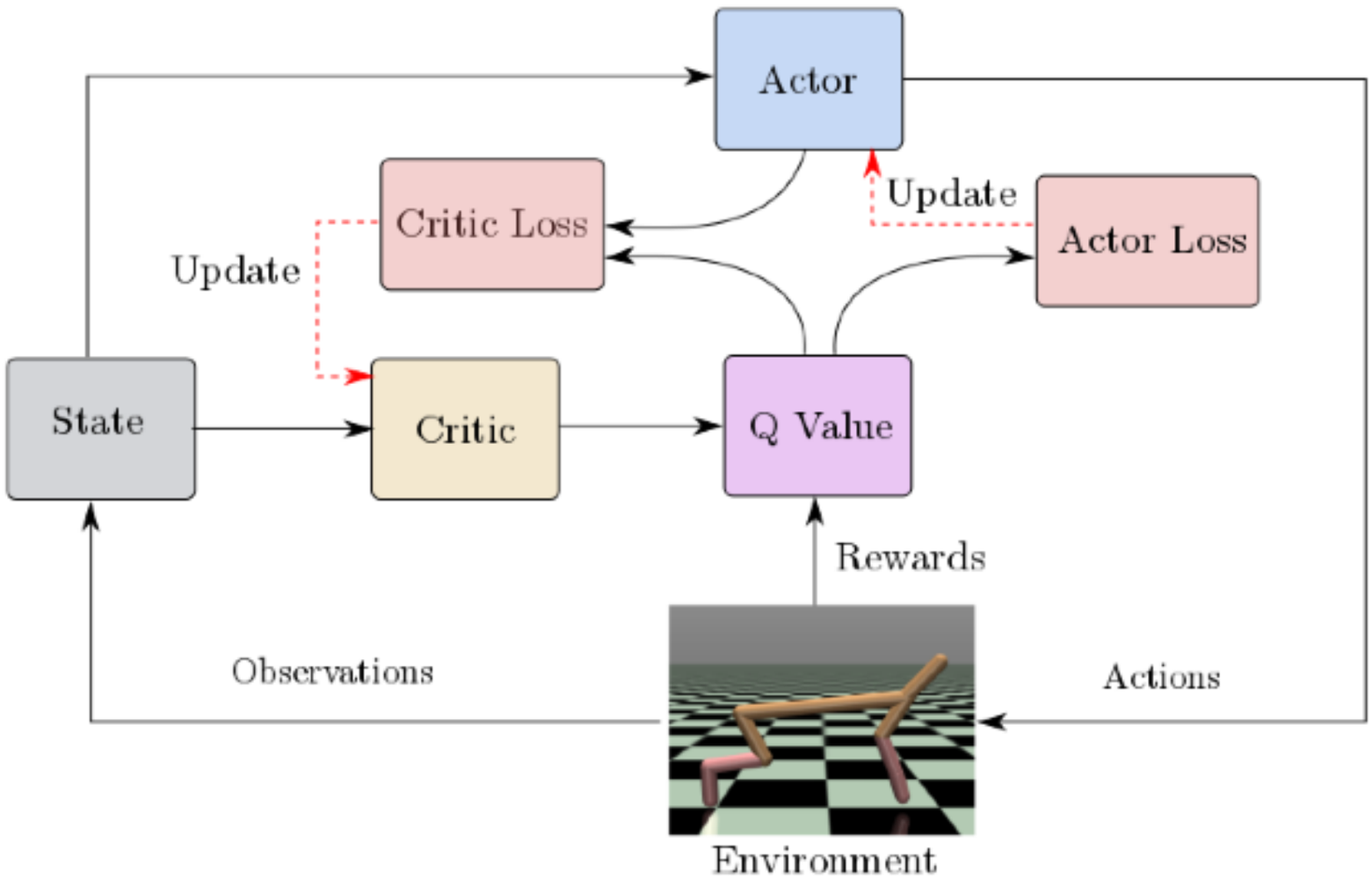
Database Generation with Curiosity Driven Exploration

- Deep Deterministic Policy Gradients (DDPG)



DDPG

- Actor tries to approximate the best policy which maps a state to optimal action

$$\mu(\theta^{\mu}) : s_t \rightarrow a_t$$

- Critic tries to approximate the predict the correct Q value

$$Qc(\theta^{Qc}) : s_t, a_t \rightarrow Q$$

- Critic is trained to Satisfy Bellman Equation

$$L(\theta^{Qc}) = (Qc - (r_t + \gamma Q(s_{t+1}, a_{t+1})^{\pi}))^2,$$

- Actor is trained by policy gradients given by,

$$\frac{\delta Qc}{\delta \theta^{\mu}} = \frac{\delta Qc}{\delta a} \frac{\delta a}{\delta \theta^{\mu}},$$