

# RL

simulation environment

state

history of observations  $h_t = \{o_0, o_1, \dots, o_t\}$

Exp. returns w/ policy  $\pi$

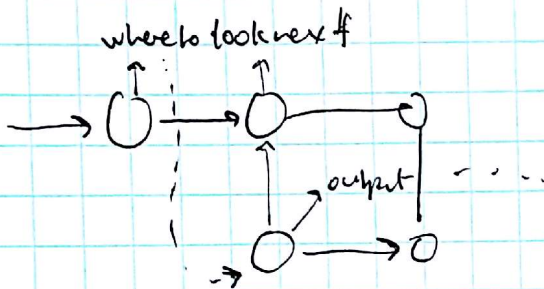
$$J^\pi(\theta) = E \left[ \sum_{t=0}^T r_t(a_t) \right]$$

reward

$$\pi_\theta(a_{0:T} | h_{0:T}) = \prod \pi_\theta(a_t | h_t)$$

policy "maps" obs. to actions  $\vec{a}$

attention direction



LSTMs

PI07 ~ pin 24 ~ CS = green

miso1 - 2-2 = pin 26 = blue<sup>11k</sup>

mosi1 - 2-3 = pin 38 ~ turquoise

sck1 - 2-1 ~ pin 13 = yellow

1 FFFF(h)

17 bit

7 bit don't care

128 KB

~~4A~~