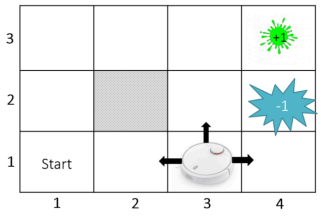
# AIoT Exercise

## Optimal Value Function $V^*$

$$V^*(s) = \max_{\pi} E[\sum_{t=1}^{H} \gamma^t R(s_t, a_t, s_{t+1}) \mid \pi,\ s_0 = s\ ]$$

= sum of discounted rewards when starting from state s and acting optimally.




- **Assumption**:
  - Actions successful probability is 0.8, $\gamma = 0.9$ , H= 100
- $V^*(4,3) = 1$
- $V^*(3,3) = +0.8 * 0.9 * V^*(4,3) + 0.1 * 0.9 * V^*(3,3) + 0.1 * 0.9 * V^*(3,2)$
- $V^*(2,3) =$
- $V^*(1,1) =$
- $V^*(4,2) =$

Exercise

Q1.: $V(2,3) = 0.8 * 0.9 * V(3,3) + 0.1 * 0.9 * V(2,4) + 0.1 * 0.9 * V(2,3)$

Q2: $V(1,1) = 0.8 * 0.9 * v(1,2) + 0.1 * 0.9 * V(2,1) + 0.1 * 0.9 * V(1,1)$