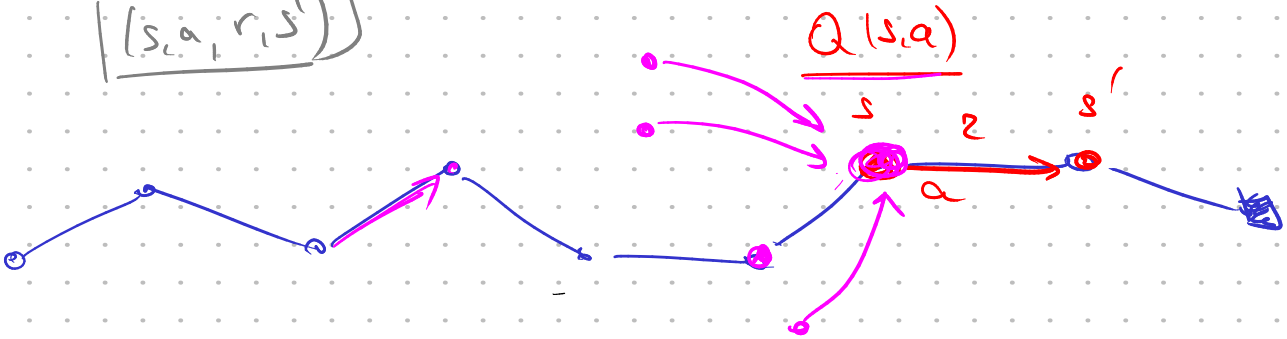
$$Q(s,a) := Q(s,a) + \alpha\left(r + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a)\right)$$

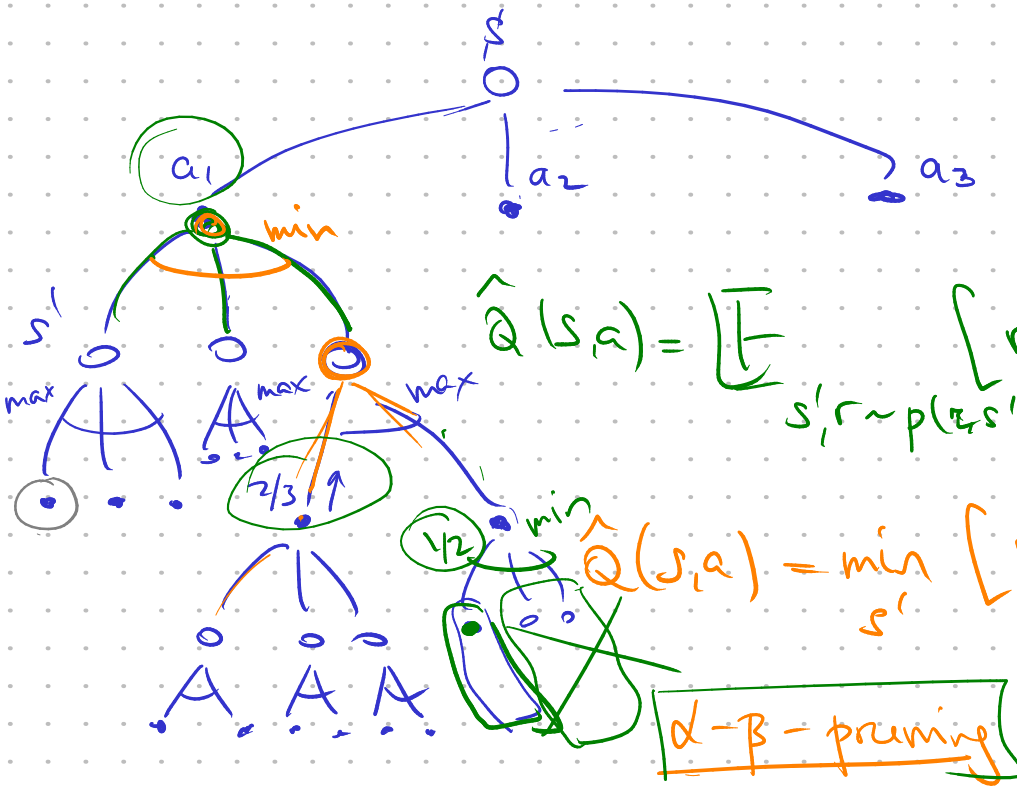
(s,a,r,s')



$V(s)$ $\quad$ $Q(s,a)$

Decision-time planning

$$\pi(s) = \operatorname{argmax}_{a'} Q(s,a')$$

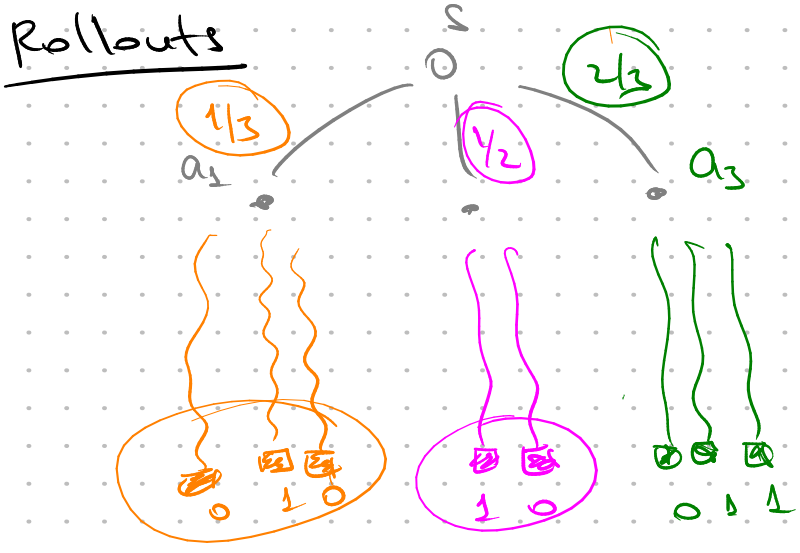


$$\hat{Q}(s,a) = \mathbb{E}_{s',r \sim p(r,s'|s,a)}\left[r + \gamma \cdot \max_{a'} Q(s',a')\right]$$

$$\hat{Q}(s,a) = \min_{s'}\left[r + \gamma \cdot \max_{a'} Q(s',a')\right]$$

α-β-pruning

Rollouts

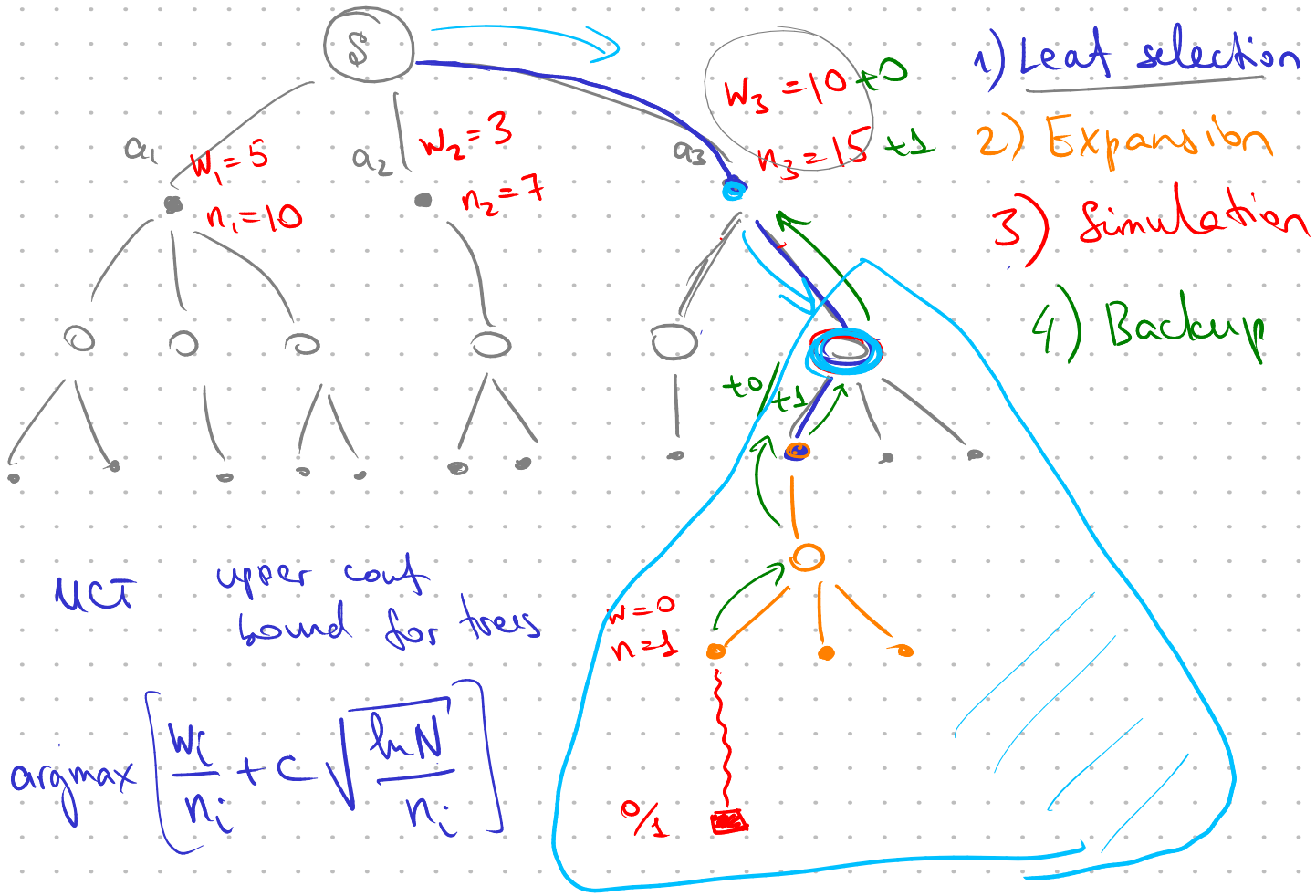


$$\hat{Q}(s,a_3) = \alpha Q(s,a_3) + (1-\alpha)\cdot \hat{R}$$

TD-Gammon

Tesauro

MCTS - Monte Carlo Tree Search



UCT upper conf bound for trees

$$\text{argmax}\left[\frac{w_i}{n_i} + c\sqrt{\frac{\ln N}{n_i}}\right]$$