Simulation (Selection: max(Q + U)

2. expand leaf and evalute: V(leaf) P(next) X(600)3. BP: N++, W+V, Q = W/N  $V(s,a)^{/2}$   $V(s,a)^{/2}$ 

L=(Z-V)- Tlogp+c1/01/2

Self-play: 
$$25,000$$
 Games

(S,  $\pi$ ,  $\approx$ )

500,000 Games (baich: 2048)  $\rightarrow$  [S