Monte Carlo Tree Search

A. Selection → **B. Expansion** → **C. Evaluation** → **D. Backpropagation**

A. Selection

In each node, select action a_k according to the **UCB** score:

$$a_k = \arg\max_a \left[Q(o,a) + c \cdot \frac{\sqrt{\sum_b N(o,b)}}{1+N(o,a)} \right]$$

The selection function is applied recursively until a **leaf** node is reached.

B. Expansion

The **leaf node** o_l is added to the search tree. Each edge (o_l, a) from the newly expanded node is initialized to:
$\{N(o_l,a) = 0, Q(o_l,a) = 0\}$

C. Evaluation

The leaf node is evaluated by running a rollout to the **end** of the game with the fast **rollout policy** (potentially random), then computing the winner with function r.

D. Backpropagation

Action values Q are updated to track the mean value of all evaluations $\{r\}$ in the subtree below that action. N are updated to track how many times the action has been on the whole MCTS simulations.
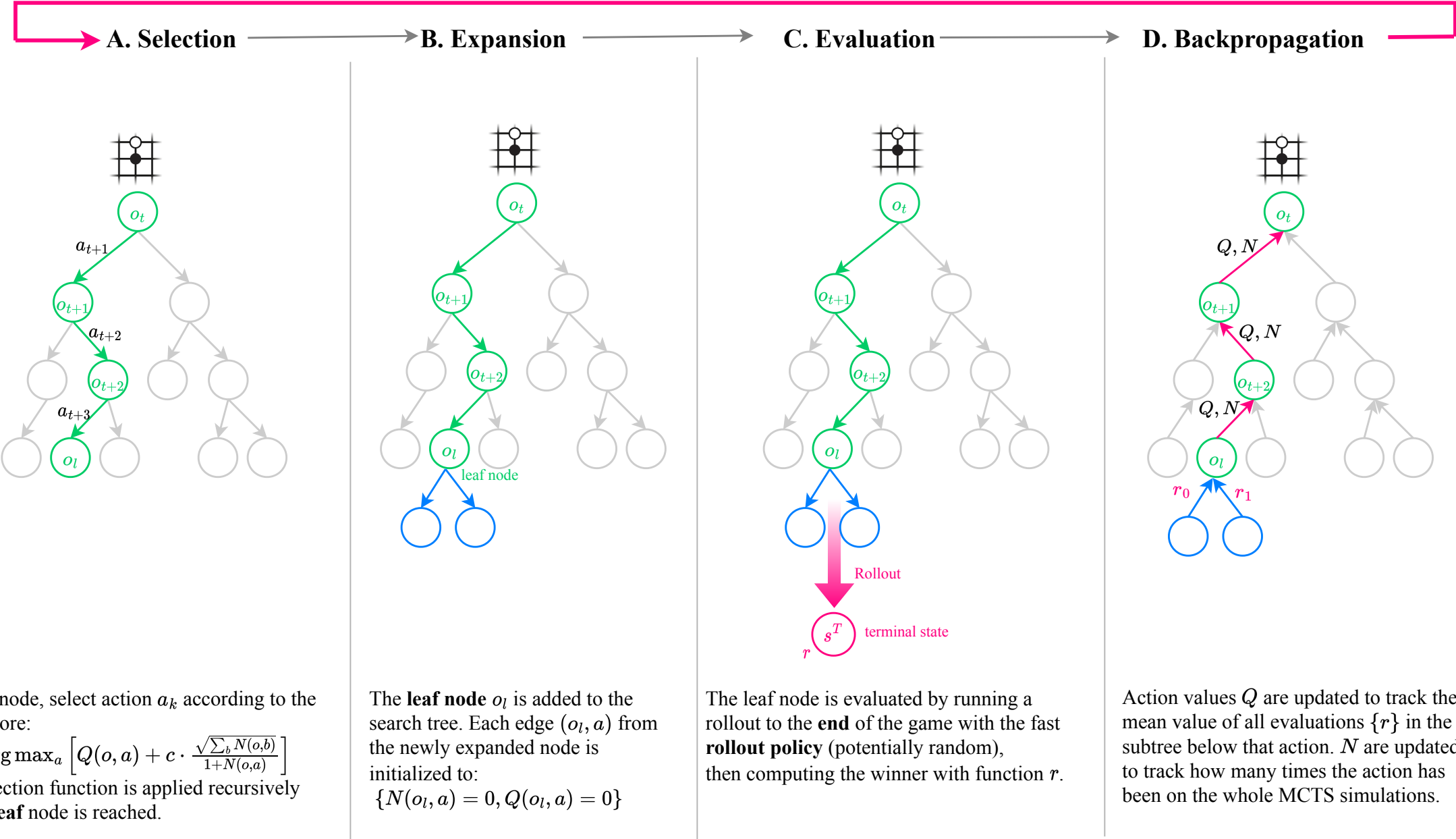
NOTE：The MCTS is divided into **four** stages, and **repeated for a number of simulations**.
Once the search is complete, we select the action with the **highest** number of visit.