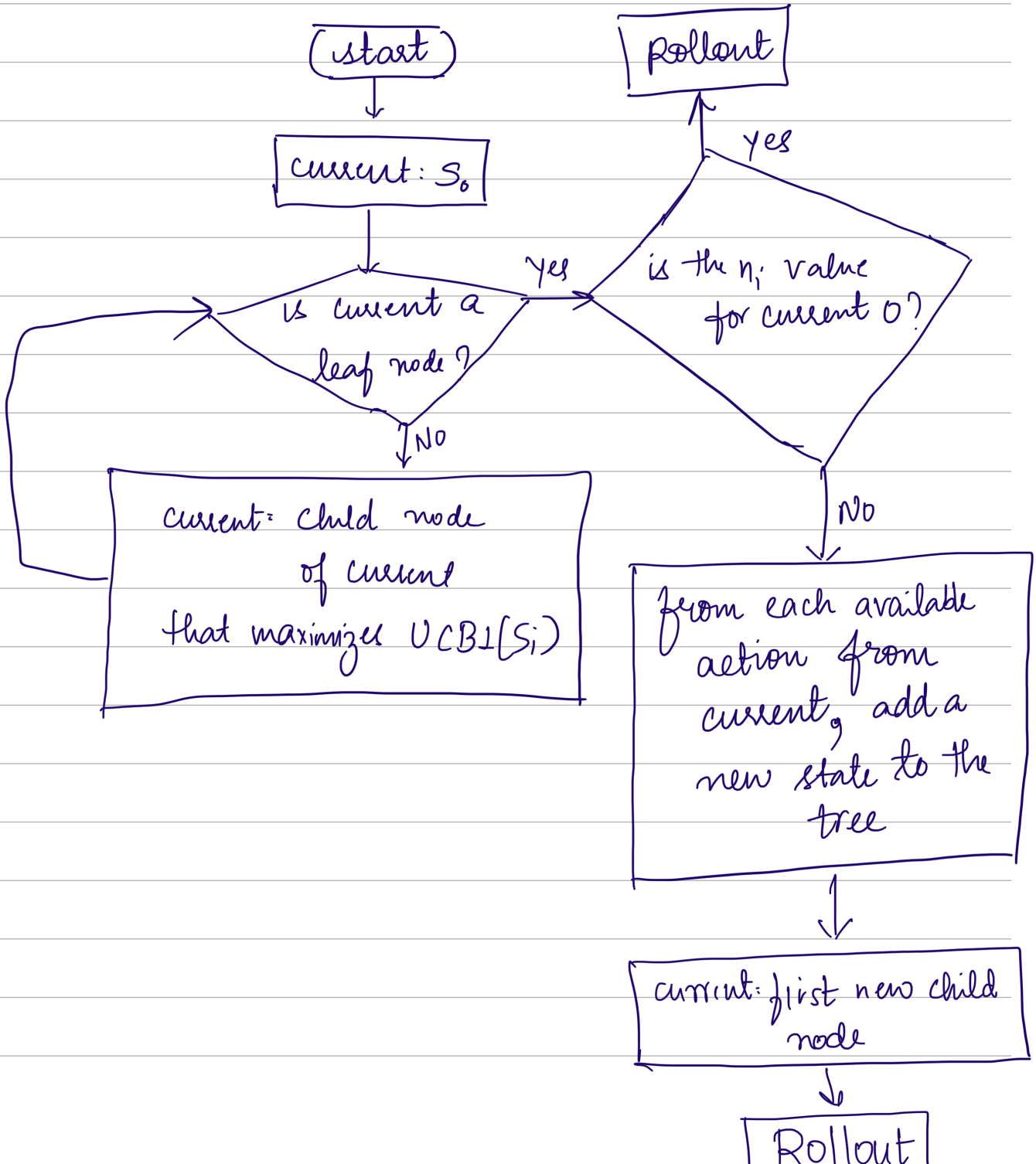
Monte-Carlo Tree Search

(selection)

1. Tree Traversal $UCB(S_i) = \overline{V}_i + C\sqrt{\frac{\ln N}{n_i}}$, $C = 2$
2. Node Expansion
3. Rollout (random simulation)
4. Backpropogation

Alg for Tree Traversal, Node Expansion



Rollout (S_i)

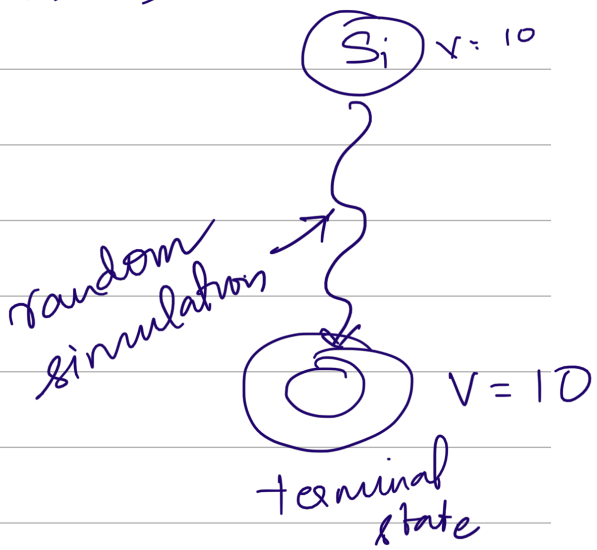
loop forever:

if S_i is a terminal state:

return value (S_i)

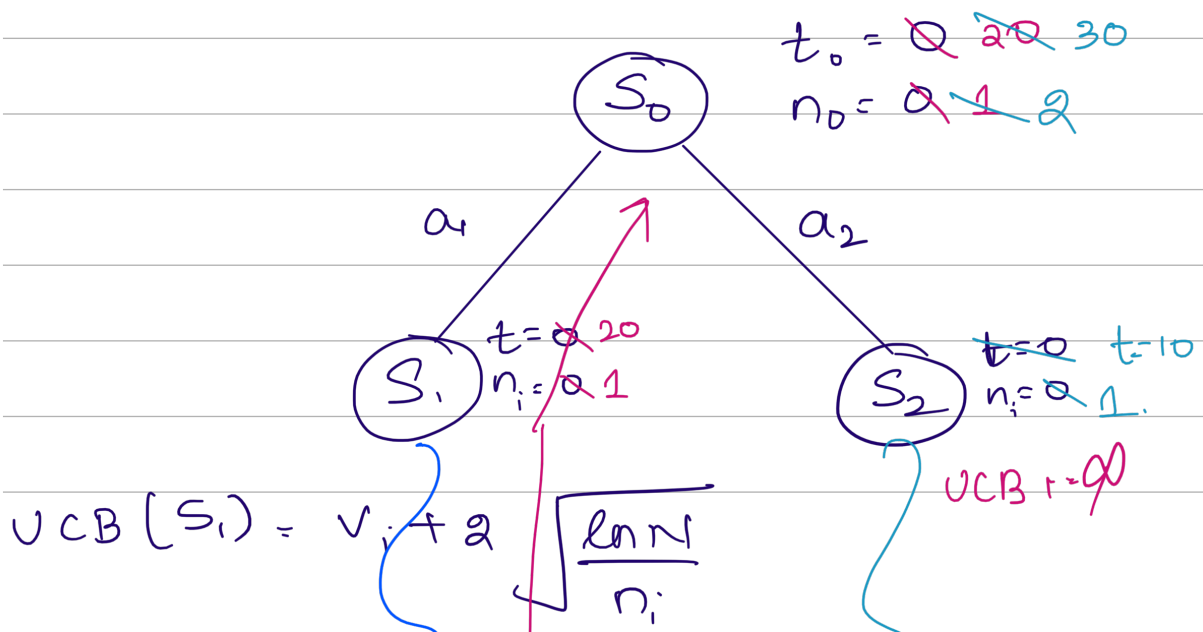
A_i = random (available_actions (S_i))

S_i = simulate (A_i, S_i)



Worked example:

$$UCB1(S_i) = \underbrace{\bar{X}_i}_{\text{exploitation term}} + \underbrace{2\sqrt{\frac{\ln N}{n_i}}}_{\text{Exploration term}}$$



not visited so:- $2\sqrt{\frac{\ln N}{n_i}} = \infty$

Backpropogate

visit S_1 is $n_1 = 0$? yes then rollout.

Then forget about rollout

S_1

$20 + 2\sqrt{\frac{\ln 1}{1}}$

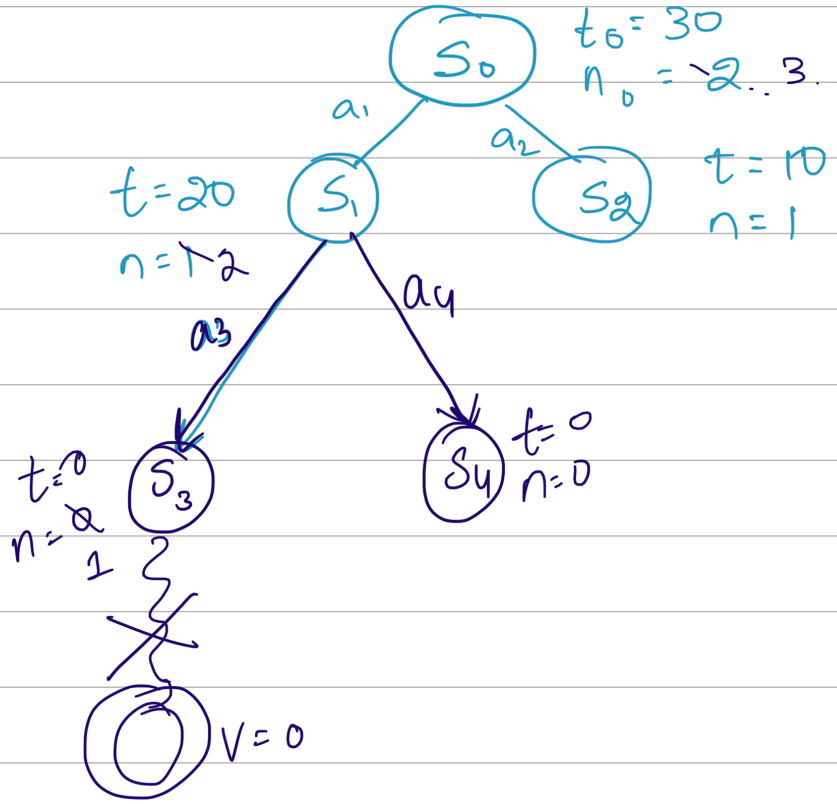
3rd iteration

S_1 $20 + 2\sqrt{\frac{\ln 2}{1}} = 21.67$

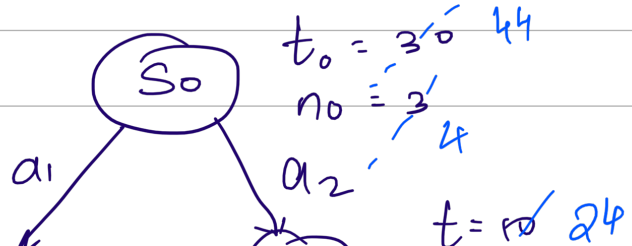
S_2 $10 + 2\sqrt{\frac{\ln 2}{1}} = 11.67$

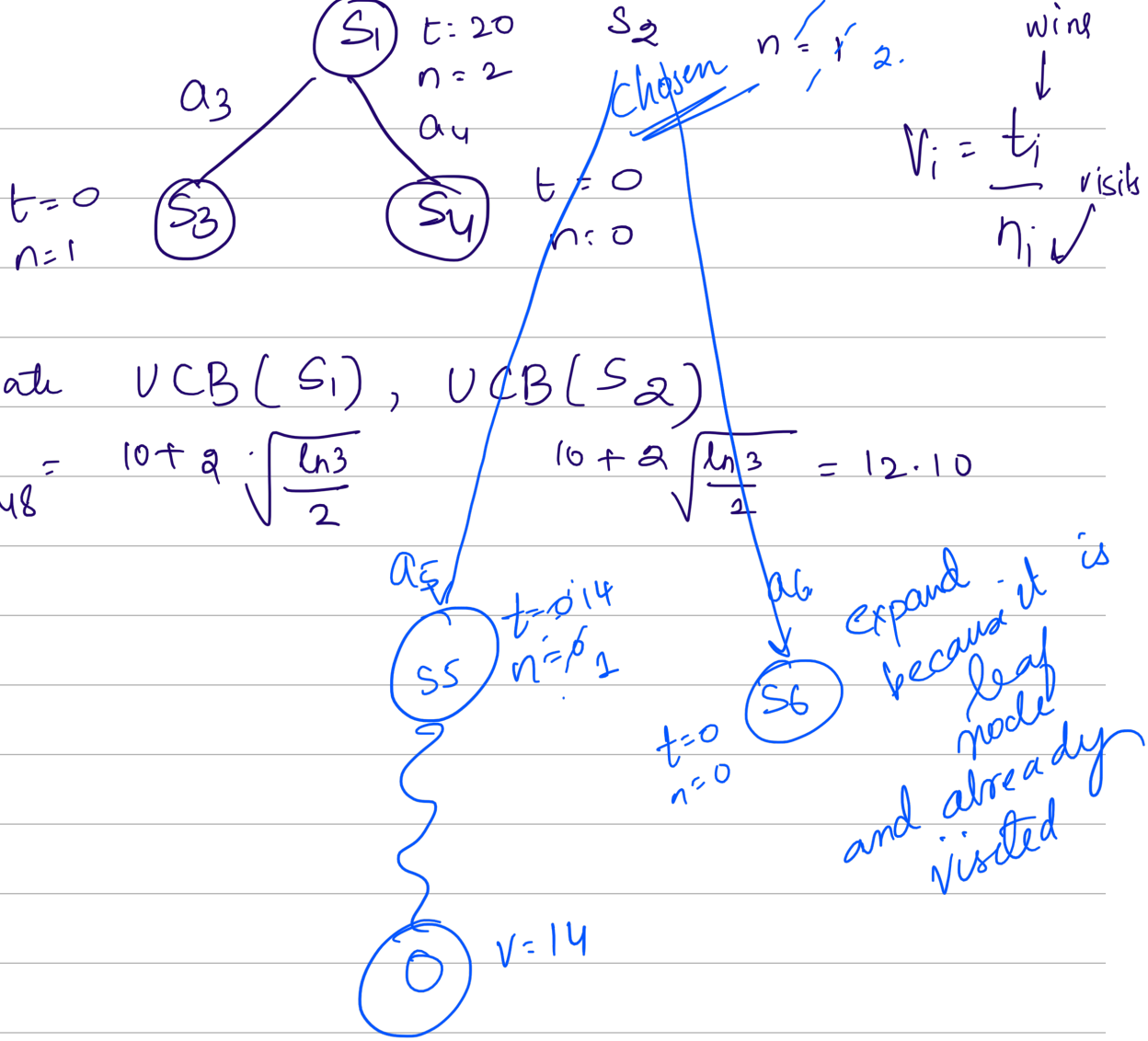
Max (21.67, 11.67) = 21.67

S_1



4th iteration





Interpretation after 4 iterations:

a_2 is better action to take because it has a higher average value