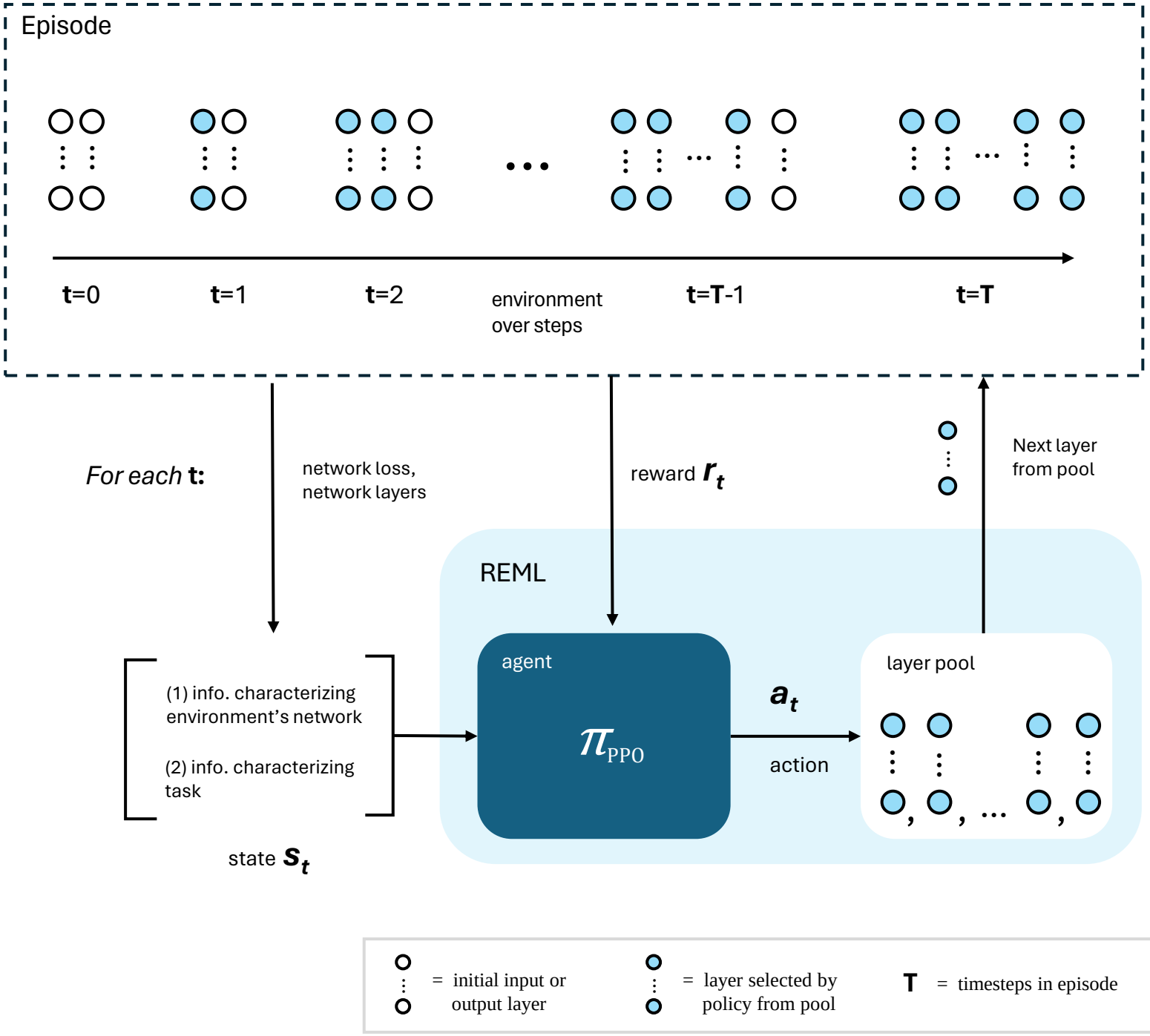
Episode
t=0
t=1
t=2
environment
over steps
t=T-1
t=T
For each t:
network loss,
network layers
reward r_t
Next layer
from pool
REML
agent
π_{PPO}
a_t
action
layer pool
(1) info. characterizing
environment's network
(2) info. characterizing
task
state s_t
= initial input or
output layer
= layer selected by
policy from pool
T = timesteps in episode