NEURAL NETWORK SURROGATE

$$\tilde{b}: (\mu, \sigma) = \phi(b)$$
belief
$$p = P_{\theta}(\tilde{b}, \cdot) \text{ policy}$$

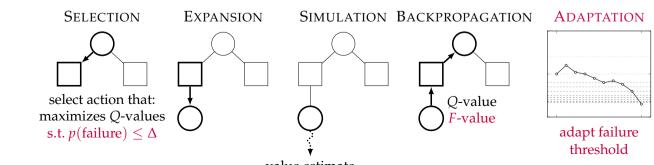
$$v = V_{\theta}(\tilde{b}) \text{ value}$$

$$p = F_{\theta}(\tilde{b}) \text{ failure probability}$$

$$(P_{\theta}, V_{\theta}, F_{\theta}) = f_{\theta}$$

OFFLINE POLICY ITERATION

Online Δ -MCTS Planning



value estimate failure probability estimate