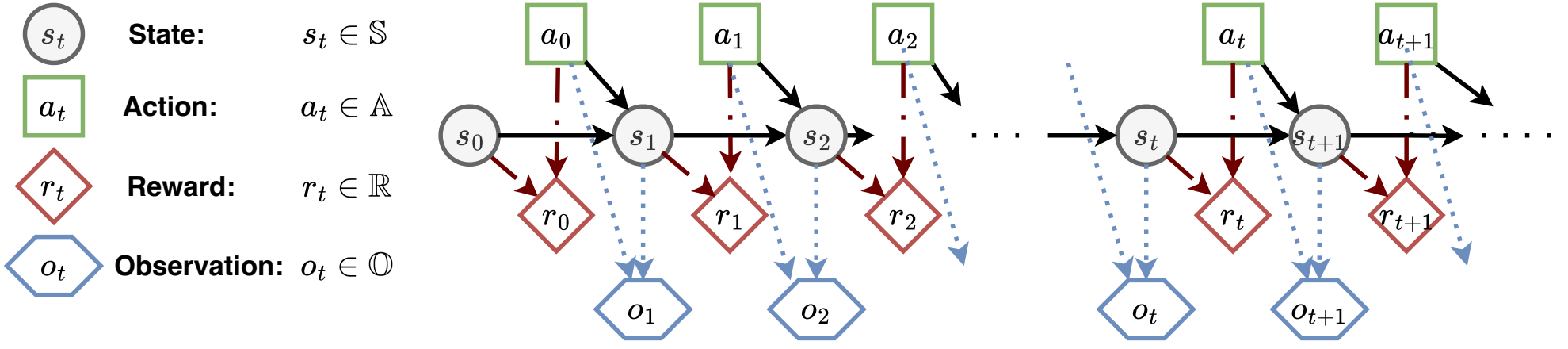

Transition Model: $\mathbf{T} = P(s_{t+1}|s_t, a_t)$

Reward Model: $\mathbf{R} = P(r_t|s_t, a_t)$

or $\mathbf{R} = r_t(s_t, a_t)$ **for deterministic reward model**

Observation Model: $\mathbf{O} = P(o_t|s_t, a_{t-1})$ **for Active Observation**

$\mathbf{O} = P(o_t|s_t)$ **for Passive Observation**