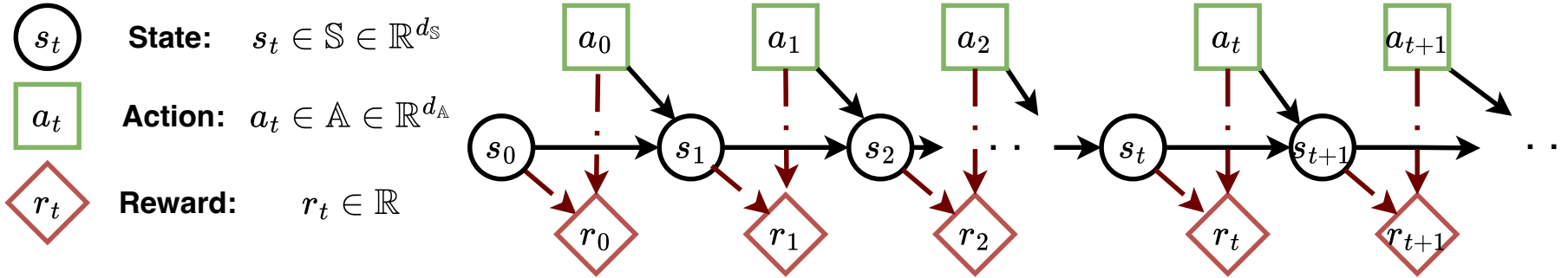
s_t
State: $s_t \in \mathbb{S} \in \mathbb{R}^{d_{\mathbb{S}}}$
a_t
Action: $a_t \in \mathbb{A} \in \mathbb{R}^{d_{\mathbb{A}}}$
r_t
Reward: $r_t \in \mathbb{R}$
a_0
a_1
a_2
a_t
a_{t+1}
s_0
s_1
s_2
s_t
s_{t+1}
r_0
r_1
r_2
r_t
r_{t+1}
Transition Model: $P(s_{t+1}|s_t, a_t)$
Reward Model: $P(r_t|s_t, a_t)$ for deterministic reward model $r(s_t, a_t)$