The overall transition model is the conditional distribution $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{X}_t, \mathbf{A}_t)$, which can be computed as a product of conditional probabilities from the DDN. The reward here is a single variable that depends only on the location $\mathbf{X}$ (for, say, arriving at a destination) and $\text{Charging}$, as the robot has to pay for electricity used; in this particular model, the reward doesn’t depend on the action or the outcome state.

The network in Figure 17.4 has been projected two steps into the future. Notice that the network includes nodes for the rewards for times $t$ and $t+1$, but the utility for time $t+2$. This