Unit Commitment

1. **States**(S_t)
    (a) Forecast demand ($d_n^f(t+1)$) at nth bus [Belief state]
    (b) Sequence of on/off status $[u_i(t), u_i^{old}(t)]$ of ith generator with a length of $\max(UT_i, DT_i) + 1$ [Physical state]
    (c) Power output $p_i(t)$ at ith generator [Physical state]
    (d) Voltage angle $\theta_n(t)$ at nth bus [Physical state] [UC-v1 env only]

2. **Actions** (A_t)
    (a) On/off status $u_i(t+1)$ at ith generator
    (b) Power output $p_i(t+1)$ at ith generator
    (c) Voltage angle $\theta_n(t+1)$ at nth bus (except the 1st bus) [UC-v1 env only]

3. **Transition function**
    (a) Forecast net load

$$d_n^f(t+2) = \text{forecast}(d_n^f(t+1))$$

    (b) On-off & Start-up & Shut-down status

$$u_i(t+1) = \text{Repair}(u_i(t+1))$$

$$v_i(t+1) = \max(0, u_i(t+1) - u_i(t+1))$$

$$w_i(t+1) = -\min(0, u_i(t+1) - u_i(t+1))$$

    (c) Sequence of On-off & Start-up & Shut-down status

$$\mathbf{u}_i(t+1) = \left[u_i(t+1), u_i^{old}(t)[:-1]\right]$$

$$\mathbf{v}_i(t+1) = \left[v_i(t+1), v_i^{old}(t)[:-1]\right]$$

$$\mathbf{w}_i(t+1) = \left[w_i(t+1), w_i^{old}(t)[:-1]\right]$$

    (d) Power output

$$p_i(t+1) = u_i(t+1)\text{Repair}(\text{clip}(p_i(t+1)))$$

(e) Voltage angle

$$\theta_n(t+1) = \text{clip}(\theta_n(t+1))$$

4. **Action Correction and Cost**

   (a) Minimum Up-time & Down-time Constraints are enforced by Repair($\cdot$) on u which uses the inequality $\sum_k^{UT} v(k) \leq u(t{+}1)$ and $\sum_k^{DT} w(k) \leq 1{-}u(t{+}1)$ to find out the generators that must be kept on & off. The violation results in a cost, but does not directly lead to "truncated".

   (b) Generation Limit Constraints are enforeced by a clip function and the corrected on/off status.

   (c) Ramping Constraints are evaluated through $p(t{+}1){-}p(t) \leq RUu(t){+}SUv(t{+}1)$ and $p(t){-}p(t{+}1) \leq RDu(t{+}1){+}SDw(t{+}1)$. Minimum Up-time & Down-time Constraints give a pair of must-on & must-off that does not contradict with each other. However, Ramping down & Minimum Down-time Constraints give a pair of must-on & must-off that may contradict with each other. If contradicted, then it results in a cost and "truncated" without good correction method. If not, then it results in a cost, but does not directly lead to "truncated". Ramping Constraints are then enforced by Repair($\cdot$) on p, which imposes a tighter bound.

5. **Reward**

   (a) Network Constraints are relaxed through slacks that represent overflow and underflow. The energy waste or unfulfillment is then explicitly computed using a load shedding cost in the reward.

   (b) Reserve Requirement Constraints are relaxed through slacks that represent the unfulfillment $r_i + s \geq R_i$. The intermediate variable r_i represents the reserve at ith generator. To compute r_i, first evaluate $\min(u(t{+}1)P_{max}{-}p(t{+}1), RUu(t){+}SUv(t{+}1){+}p(t){-}p(t{+}1)))$, then evaluate $\max(\cdot, 0)$. This finds the best reserve that can be achieved without violating both power limit and ramping-up constraints. The unfulfillment is then explicitly computed using a reserve penalty cost in the reward.