

Base policy

Define the base policy Π

Evaluate its value function $\{W_t^\Pi(Q_t^D) \mid \forall t, Q_t^D\}$ analytically

Scheduling with
value function
of base policy

Approximate the value function of optimal policy
by $\{W_t^\Pi(Q_t^D) \mid \forall t, Q_t^D\}$, then perform one-step policy iteration

Decompose **P2** into K sub-problems

Solve each sub-problem, then derive the proposed low-
complexity policy Ψ