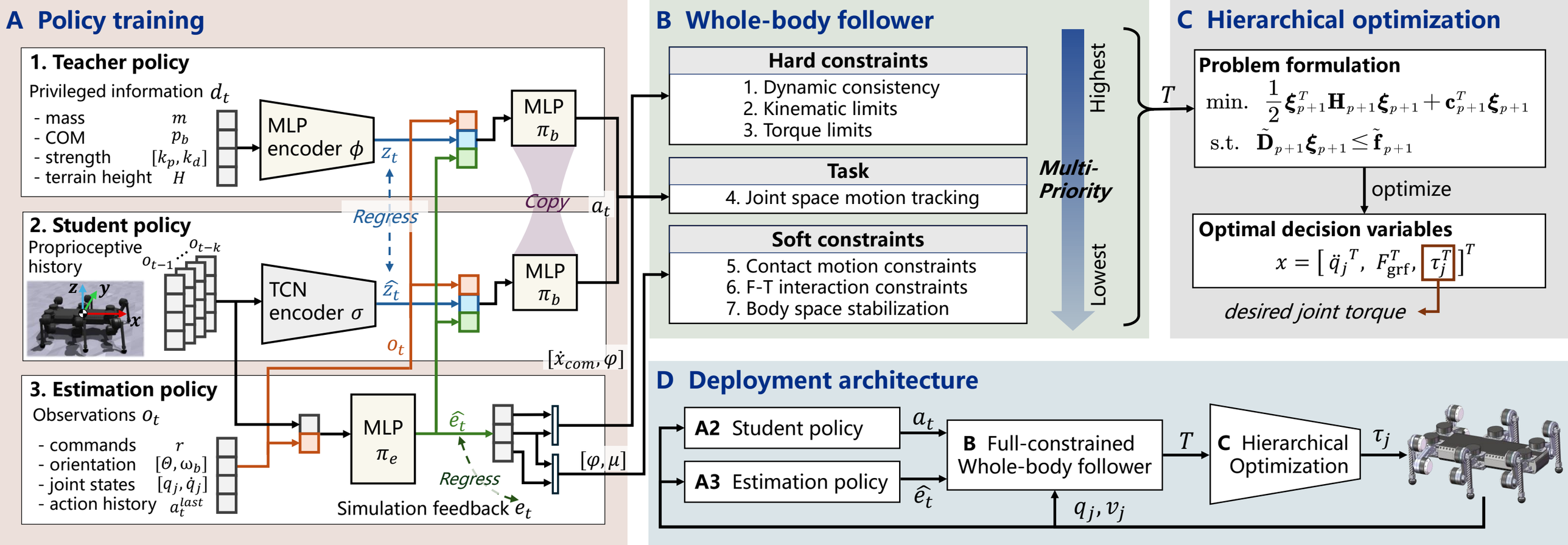
A Policy training
1. Teacher policy
Privileged information d_t
- mass m
- COM p_b
- strength $[k_p, k_d]$
- terrain height H
MLP encoder ϕ
z_t
MLP π_b
Copy
a_t
Regress
2. Student policy
Proprioceptive history
o_{t-1} ... o_{t-k}
TCN encoder σ
$\hat{z}_t$
MLP π_b
o_t
$[\dot{x}_{com}, \varphi]$
3. Estimation policy
Observations o_t
- commands r
- orientation $[\Theta, \omega_b]$
- joint states $[q_j, \dot{q}_j]$
- action history a_t^{last}
MLP π_e
$\hat{e}_t$
Regress
Simulation feedback e_t
$[\varphi, \mu]$
B Whole-body follower
Hard constraints
1. Dynamic consistency
2. Kinematic limits
3. Torque limits
Task
4. Joint space motion tracking
Soft constraints
5. Contact motion constraints
6. F-T interaction constraints
7. Body space stabilization
Multi-Priority
Highest
Lowest
T
C Hierarchical optimization
Problem formulation
$\min. \ \frac{1}{2}\boldsymbol{\xi}_{p+1}^T \mathbf{H}_{p+1}\boldsymbol{\xi}_{p+1} + \mathbf{c}_{p+1}^T\boldsymbol{\xi}_{p+1}$
$\text{s.t.} \ \tilde{\mathbf{D}}_{p+1}\boldsymbol{\xi}_{p+1} \leq \tilde{\mathbf{f}}_{p+1}$
optimize
Optimal decision variables
$x = [\ \ddot{q}_j{}^T,\ F_{grf}^T,\ \tau_j^T\ ]^T$
desired joint torque
D Deployment architecture
A2 Student policy
A3 Estimation policy
a_t
$\hat{e}_t$
B Full-constrained Whole-body follower
T
C Hierarchical Optimization
τ_j
q_j, v_j