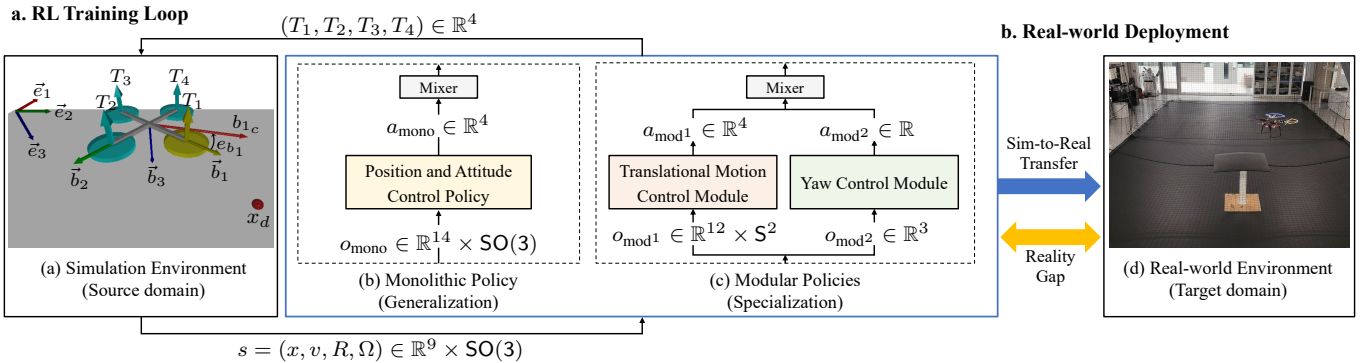

Fig. 2. A schematic overview of the system. **a.**, During training, we train RL policies for quadrotor low-level control tasks in simulation. (a) A custom simulator serves as a training environment, providing full access to the quadrotor's dynamics and state. (b) A monolithic end-to-end policy directly outputs total thrust f and moments M. (c) Two specialized modules independently control translational and yaw motions, each selecting the optimal action based on its local observations. **b.**, When transferring trained policies from simulation to the physical world, the sim-to-real gap arises from mismatches between simulation and reality. To bridge this gap, domain randomization is applied during the training phase. (d) An indoor flight test facility at the Flight Dynamics and Control Lab, GWU for real-world deployment. A supplementary video of the RL training and real-world experiments is available at https://youtu.be/TGBQTuKpbAw.