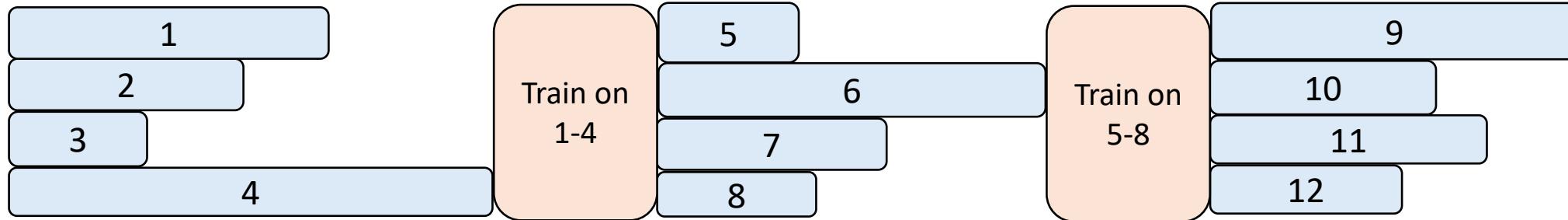
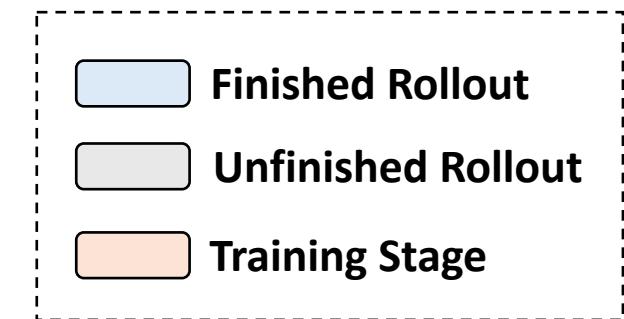
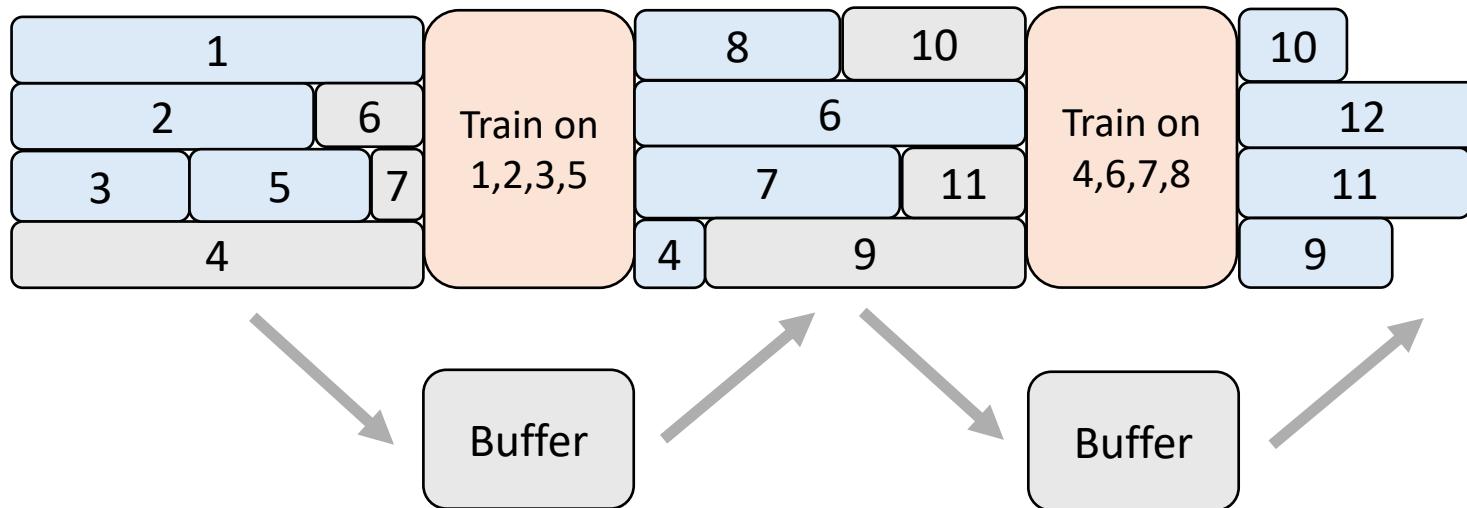


(a) Standard RL Training



(b) PRIS-augmented RL Training



Compute log prob on 4,6,7

Compute log prob on 9,10,11