

High task variance ($\mathcal{A}/g^{(1)}$ applied)

Accumulated reward

1.5
1.0
0.5
0.0
-0.5

0

1

2

3

4

Number of policy adaptation steps



BO-MRL (ours)



Random initialization



Optimal task-specific policies