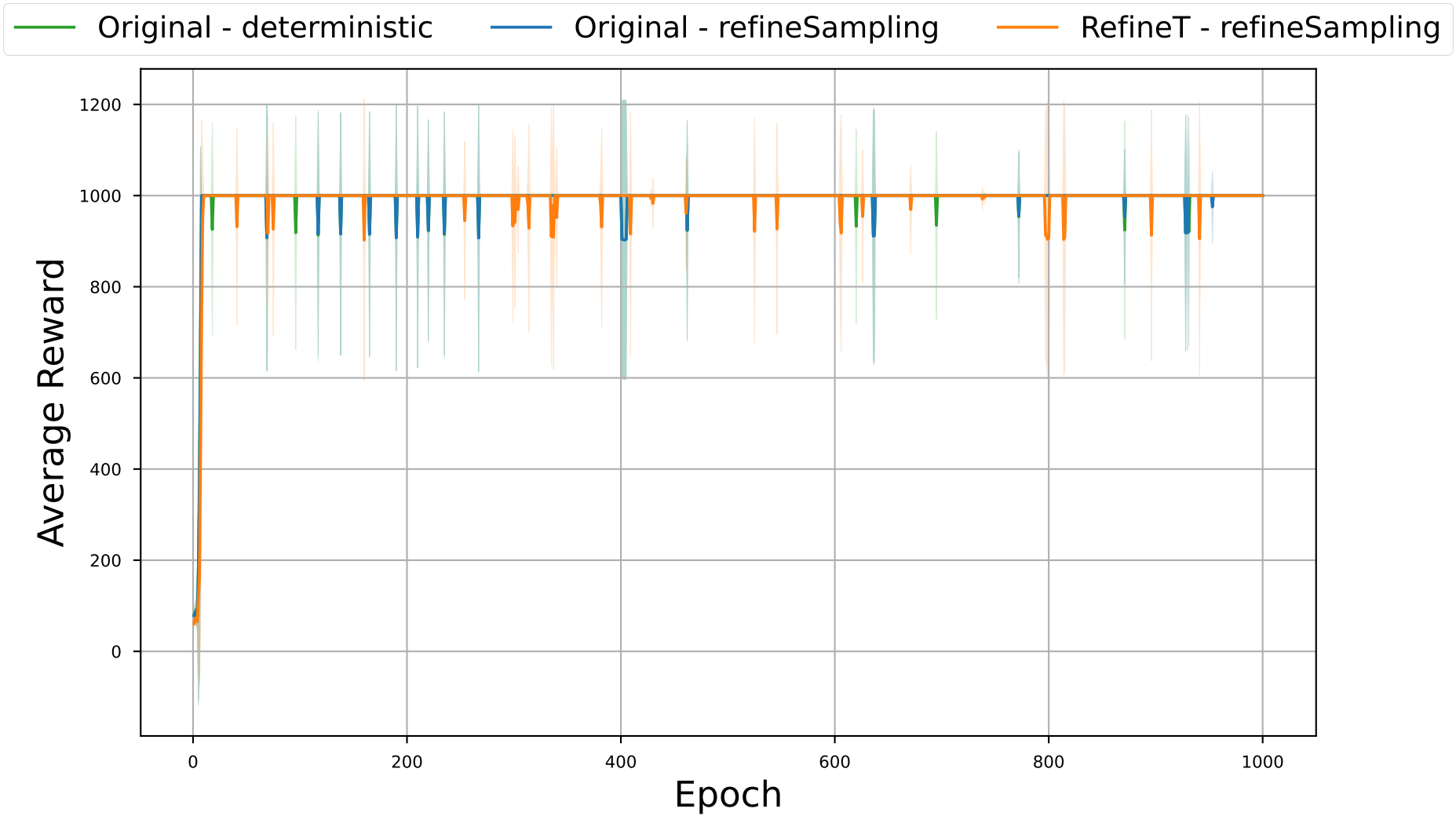
Original - deterministic
Original - refineSampling
RefineT - refineSampling
Average Reward
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
Epoch