REINFORCE P-T vs. DQN P-T
$\delta d = 0.2$ $\varepsilon d = 0.4$

- REINFORCE P-T
- DQN P-T

performance vs. context inference error