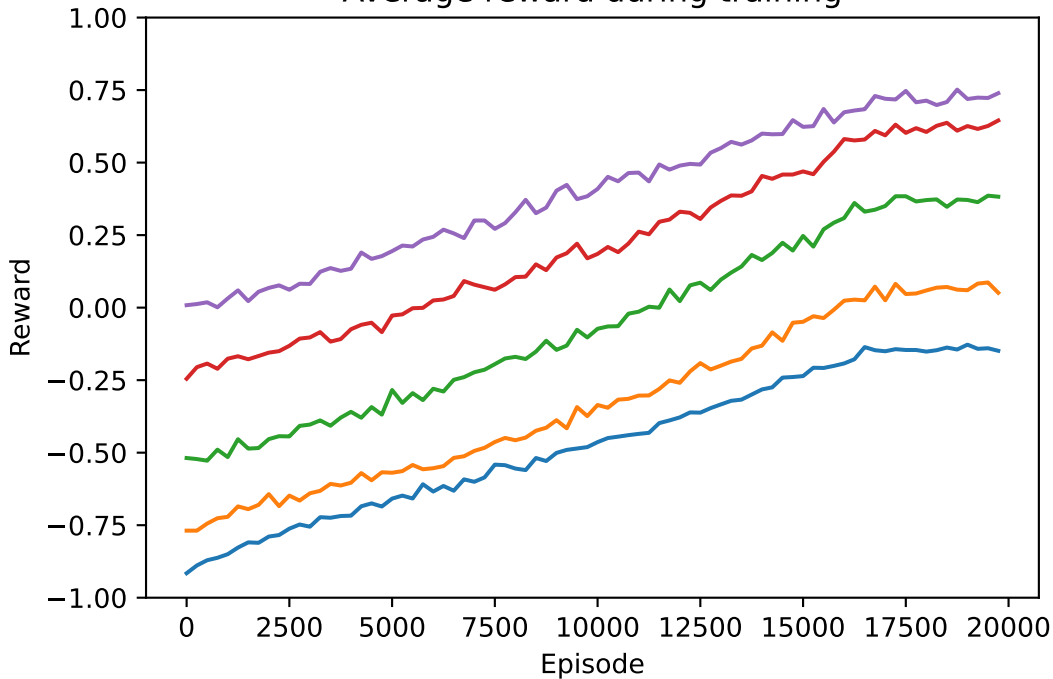


Average reward during training



$\epsilon_{opt} = 0.0$ $\epsilon_{opt} = 0.5$ $\epsilon_{opt} = 0.8$ $\epsilon_{opt} = 1.0$
 $\epsilon_{opt} = 0.2$