

# Monothread CPU program

Sync. CPU → GPU data transfer  
(default stream)

Async. GPU parallel computation

Sync. internode CPU comms

Sync. GPU → CPU data transfer

# CPU

# GPU

MPI

comms

waiting

GPU  
thread  
exec.