Choose experts with gating function, locate workers using DHT

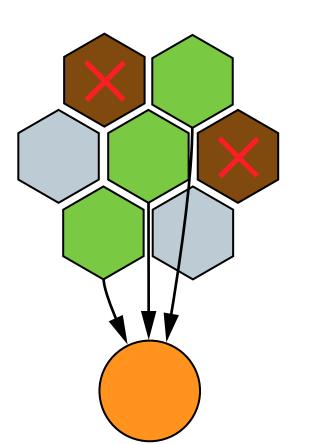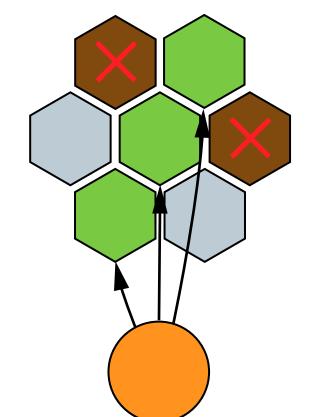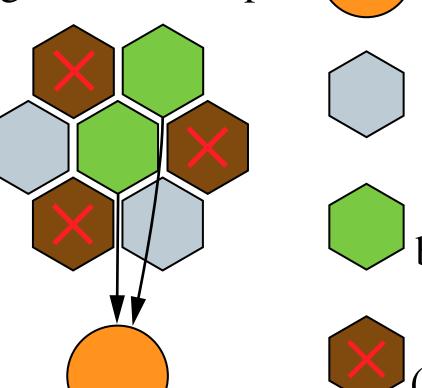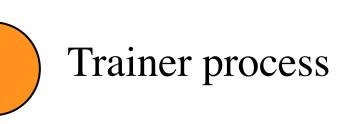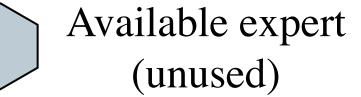Send inputs and execute a forward pass

Aggregate outputs of responding experts

On backward pass, send inputs and gradients

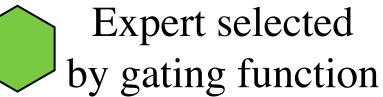Update parameters of responding experts, get gradient for input

DHT request

Data transfer

Trainer process

Available expert (unused)

Expert selected by gating function

Failed expert (e.g. disconnected)