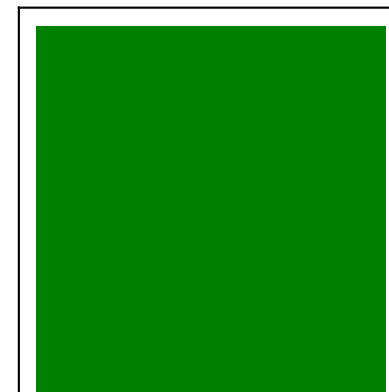
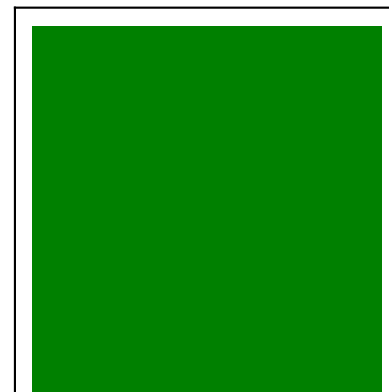
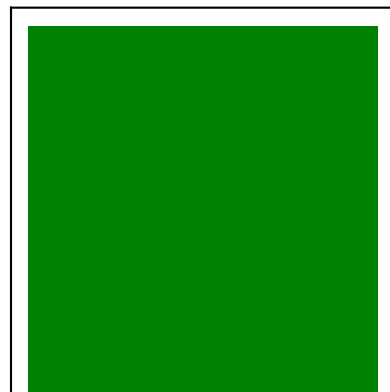
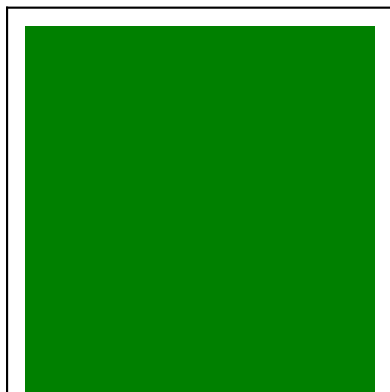
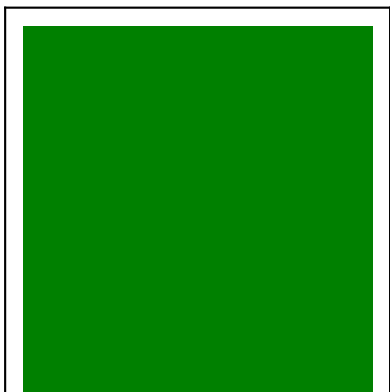
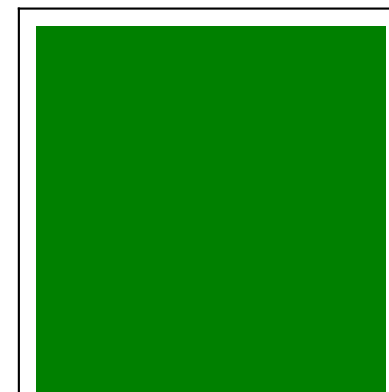
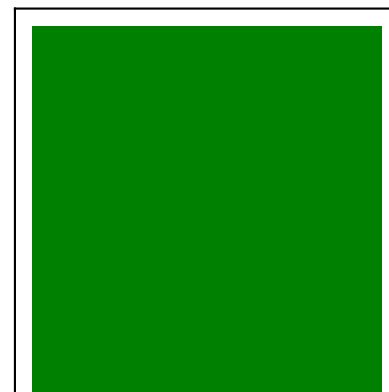
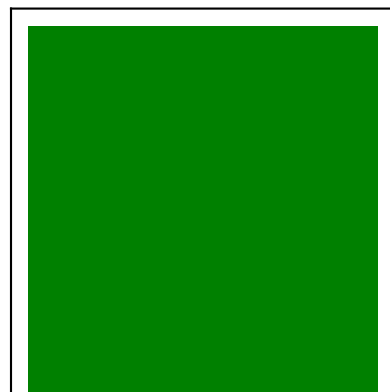
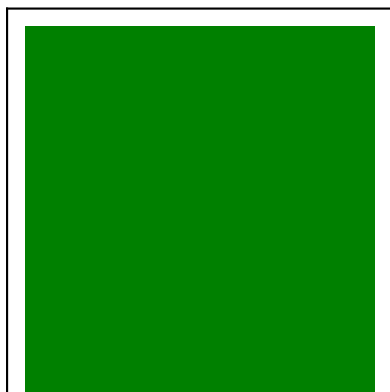
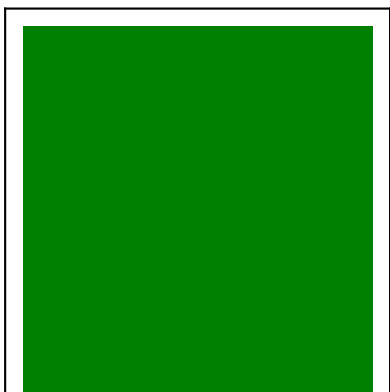


Latency Breakdown (HBM, & NVLink)

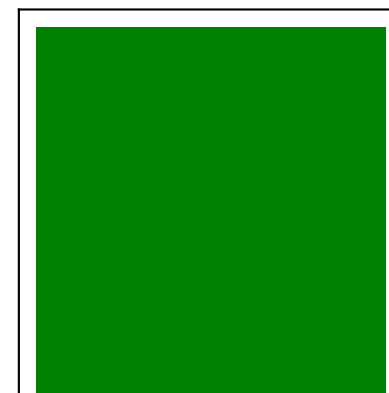
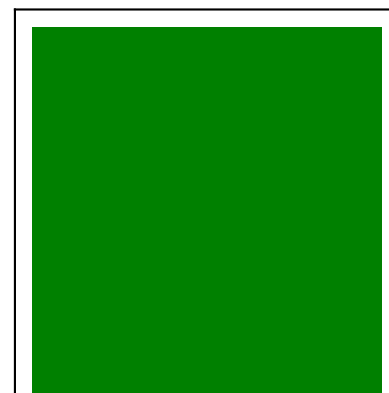
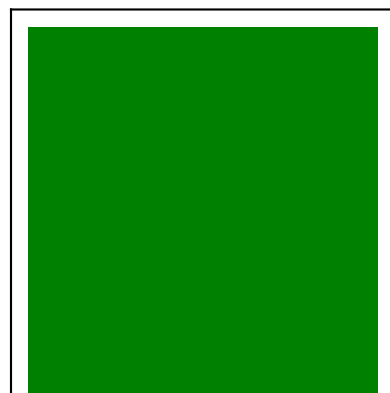
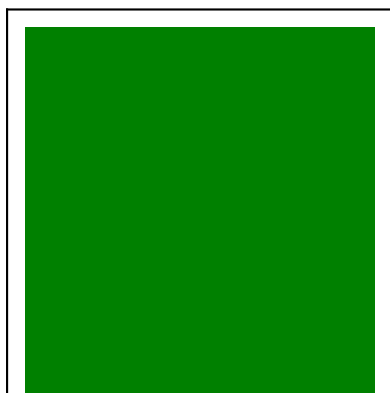
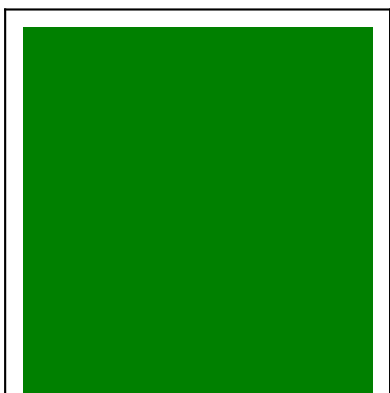
GPU



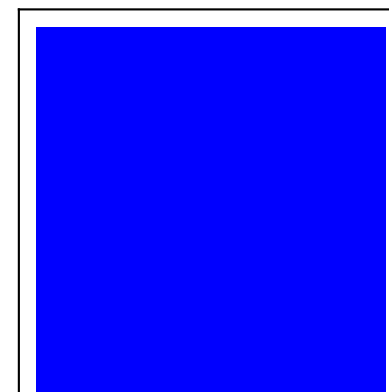
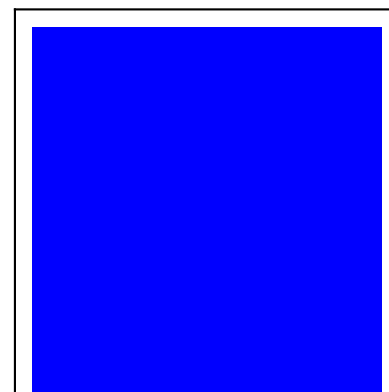
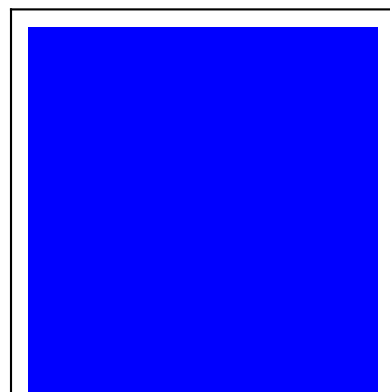
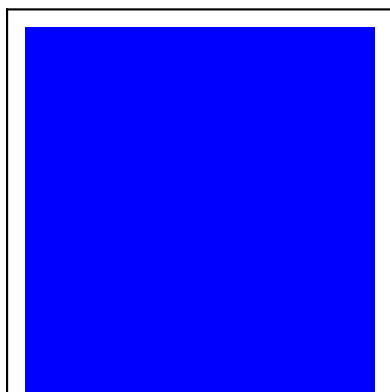
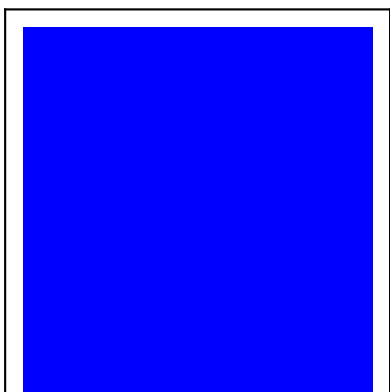
TPU



RDU



WSE



2D Torus

Dragonfly

3D Torus

DGX-1

DGX-2