

論文読み

Twitter で , @ogawa_tter さんに教えてもらったので読んでみた
Evaluating On-Node GPU Interconnects for Deep Learning Workloads
<http://www.dcs.warwick.ac.uk/pmbs/pmbs/PMBS/papers/paper1.pdf>

NVLink は速いけどワークロードによっては (ResNet みたいな) ,
PCIe 接続な GPU サーバー (GX8) でも結構いい速度でるよ , という話 .
DGX-1 の 8-GPU の NVLink での同期が遅いこと ,
メッセージサイズが小さい段階で PCIe が早めにサチること , が鍵なのかな .