ISCAS 中国科学院软件研究所学术年会’2023
暨计算机科学国家重点实验室开放周

学术论文

异构多协处理器环境下图计算与图神经网络系统

张珩，宦成颖，武延军
智能软件研究中心

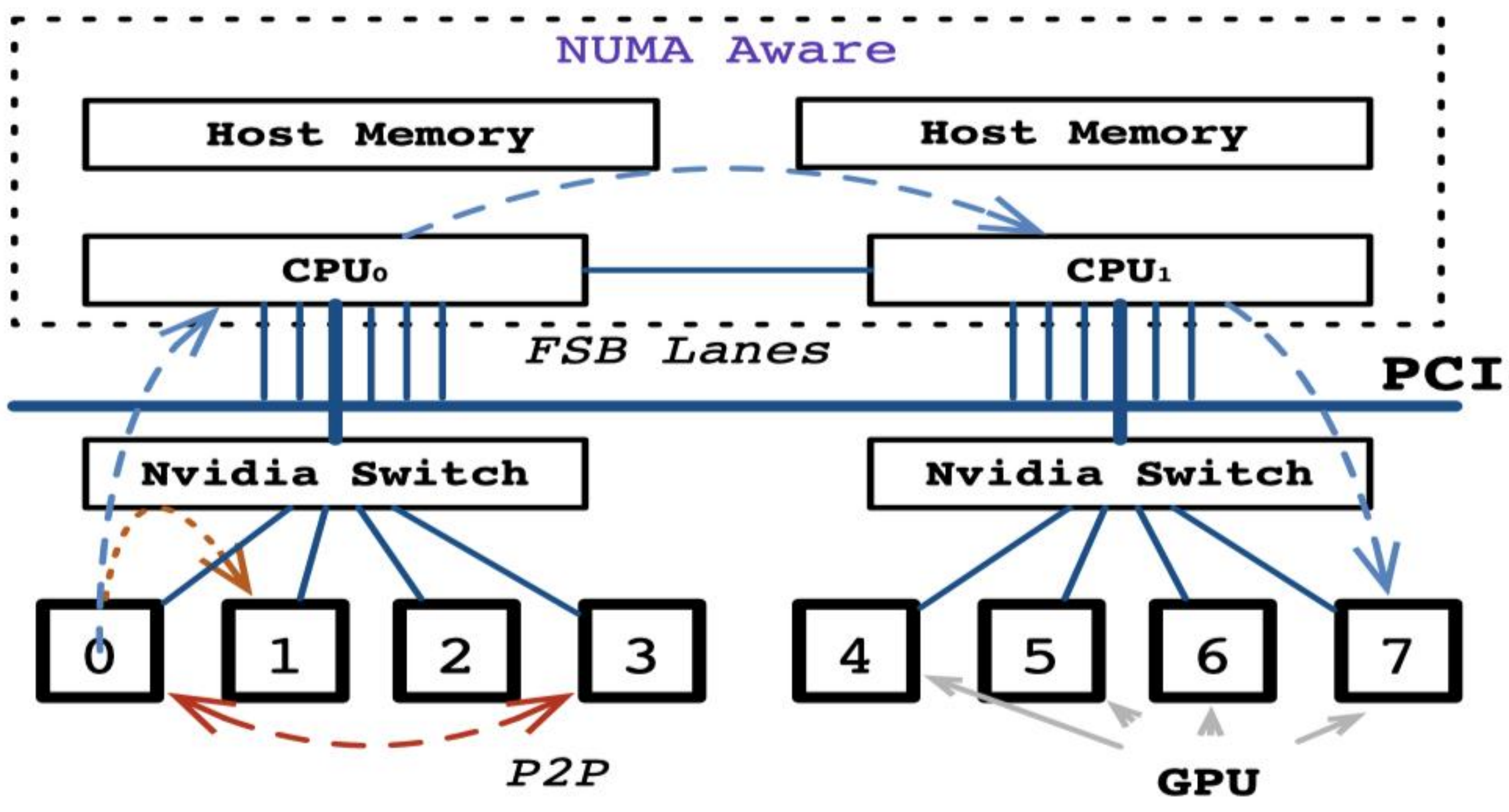
ChattyGraph: 面向异构多协处理器的高可扩展图计算系统. 软件学报, 2023, 34(4).
T-GCN: A Sampling Based Streaming Graph Neural Network System with Hybrid Architecture. PACT 2022.ACM, New York, NY, USA, 69–82.

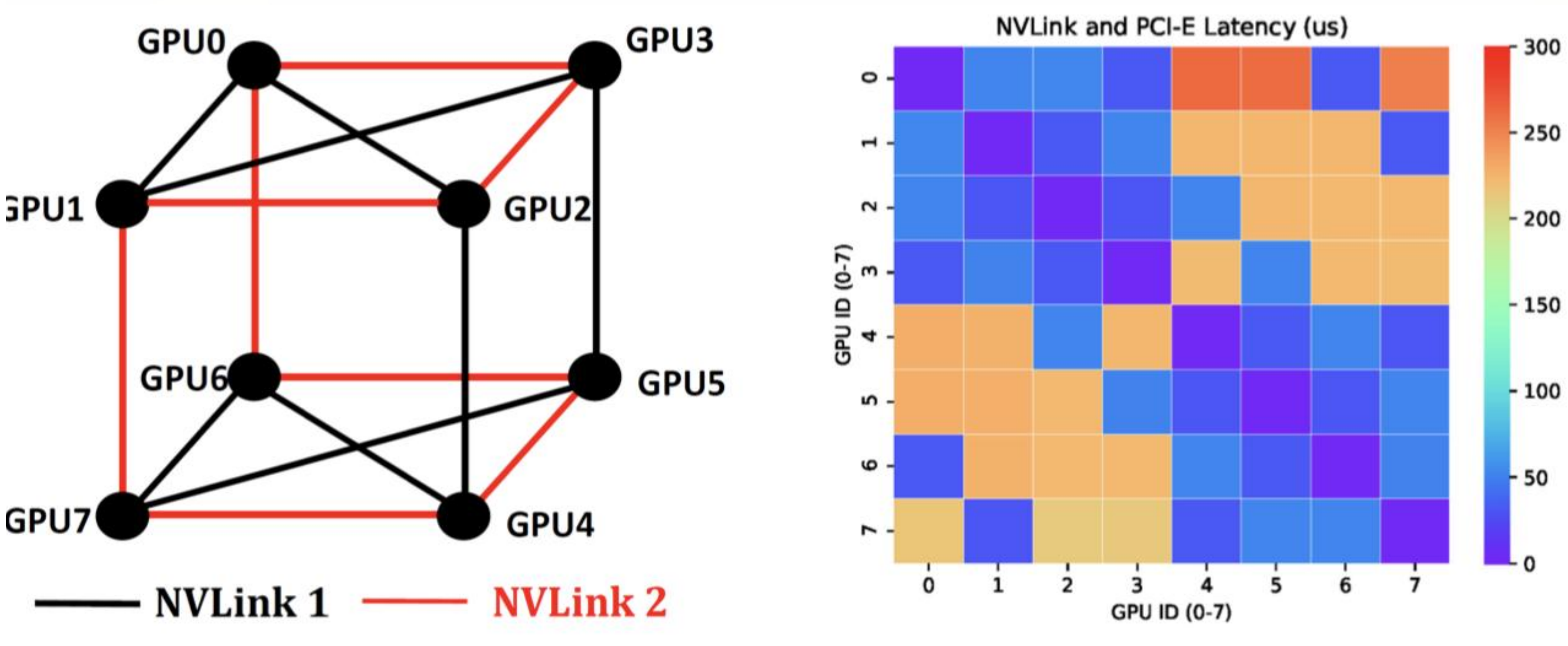
联系方式： 张珩, 15652191318, zhangheng17@iscas.ac.cn

现代多协处理器环境架构：非中心化设备互联

Multi-GPU大规模数据处理系统的性能量化

- ❖NVLink、CXL：现代链路的计算框架
- ❖PCI-E：中心化经典框架





(a) NVIDIA DGX 的拓扑结构

(b) GPU 间通信带宽热力图

1）物理链路评估：PCI-E、NVLink V1&V2

表 1 GPU 间链路互联特性.内存访问带宽(MB/s)和延迟(ms)

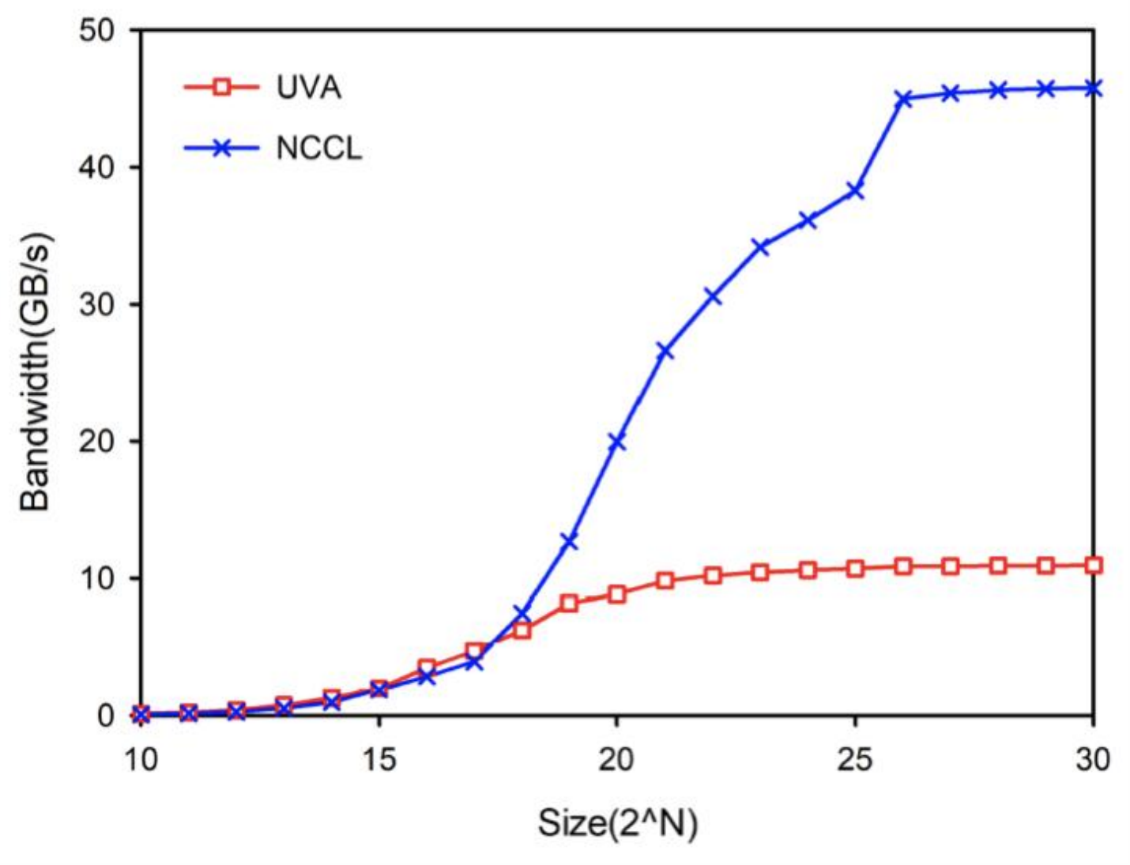
设备链路	带宽(MB/s)	0-hop (ms)	1-hop (ms)	2-hop (ms)
PCI-E	2872.111	0.348176	0.686045	0.732432
NVLink-V1	17196.9	0.05815	0.117411	0.176563
NVLink-V2	27228.67	0.036726	0.07807	0.11305

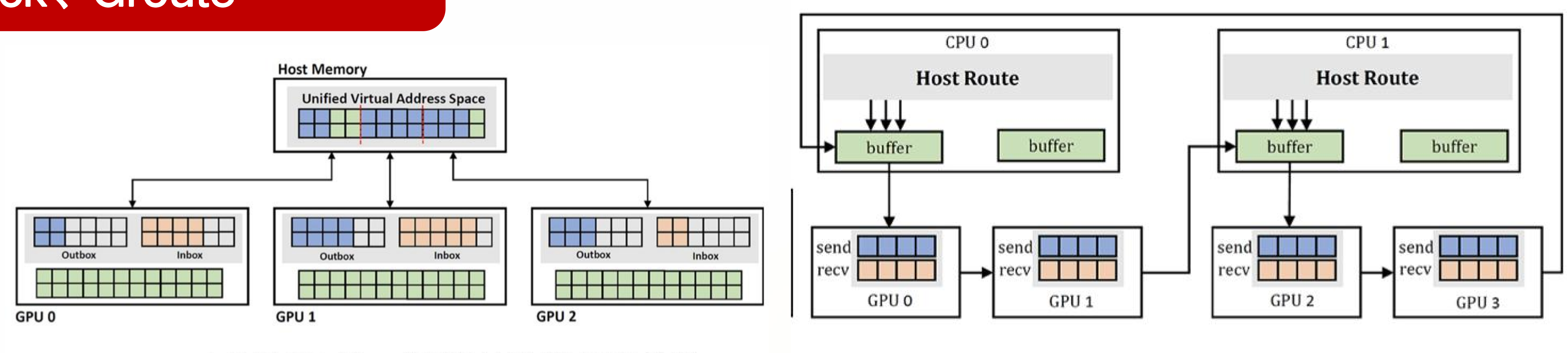
2）数据通信方式：显式接口、RingReduce、UVM

表 2 点对点通信在传输不同大小数据时通信延迟(ms)

设备间数据传输大小 Size(KB)	设备间零拷贝 Zero-Copy（ms）	同步式设备间内存访问 MemCopy（ms）	异步式设备间内存拷贝 MemCpyAsync（ms）
4	0.017	0.040	0.025
16	0.037	0.061	0.050
64	0.118	0.136	0.124
256	0.621	0.405	0.372
1024	3.317	1.497	1.419
4096	25.272	5.391	5.307
16384	174.579	27.144	28.302

相关工作分析：WSVR、Gunrock、Groute

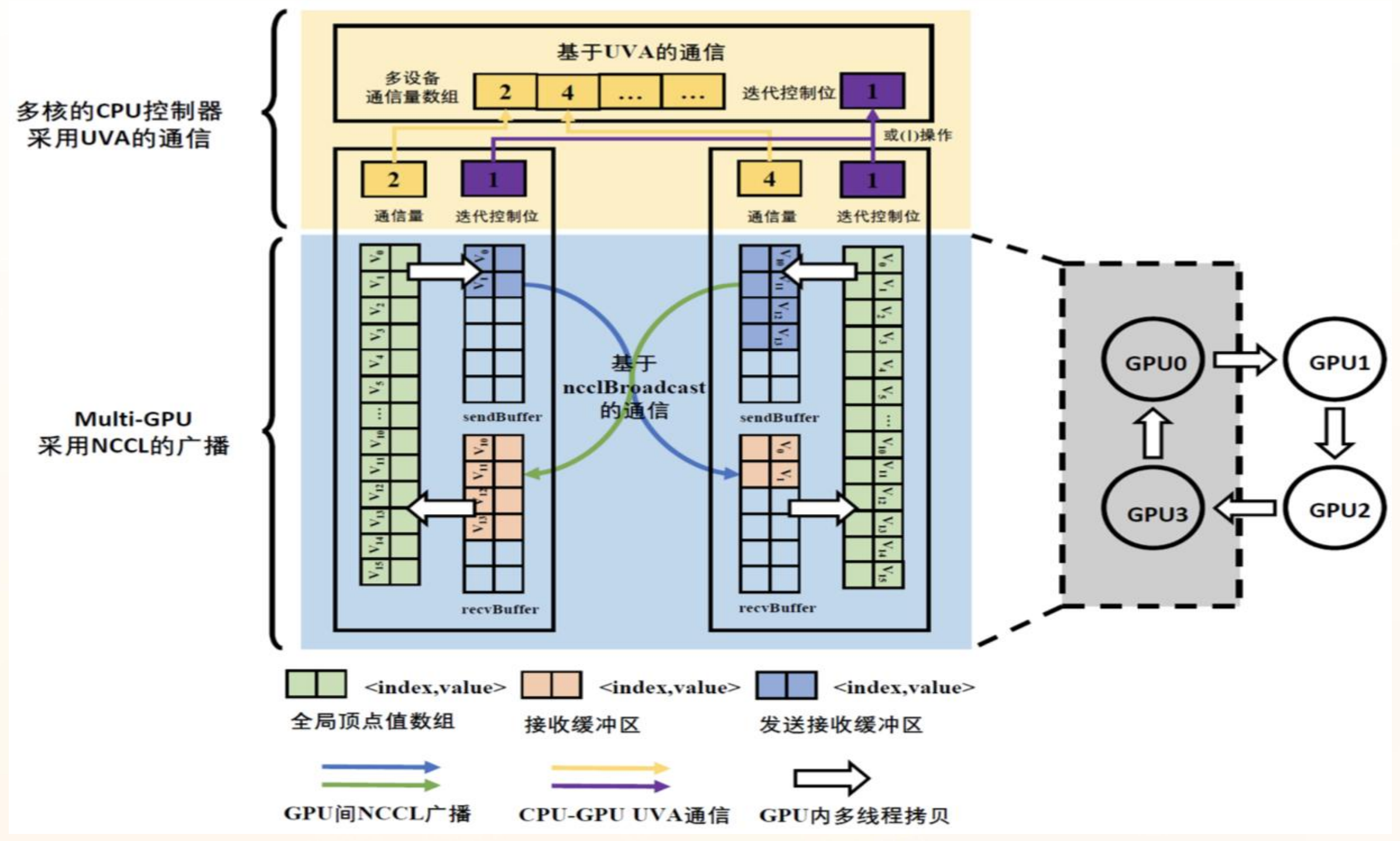


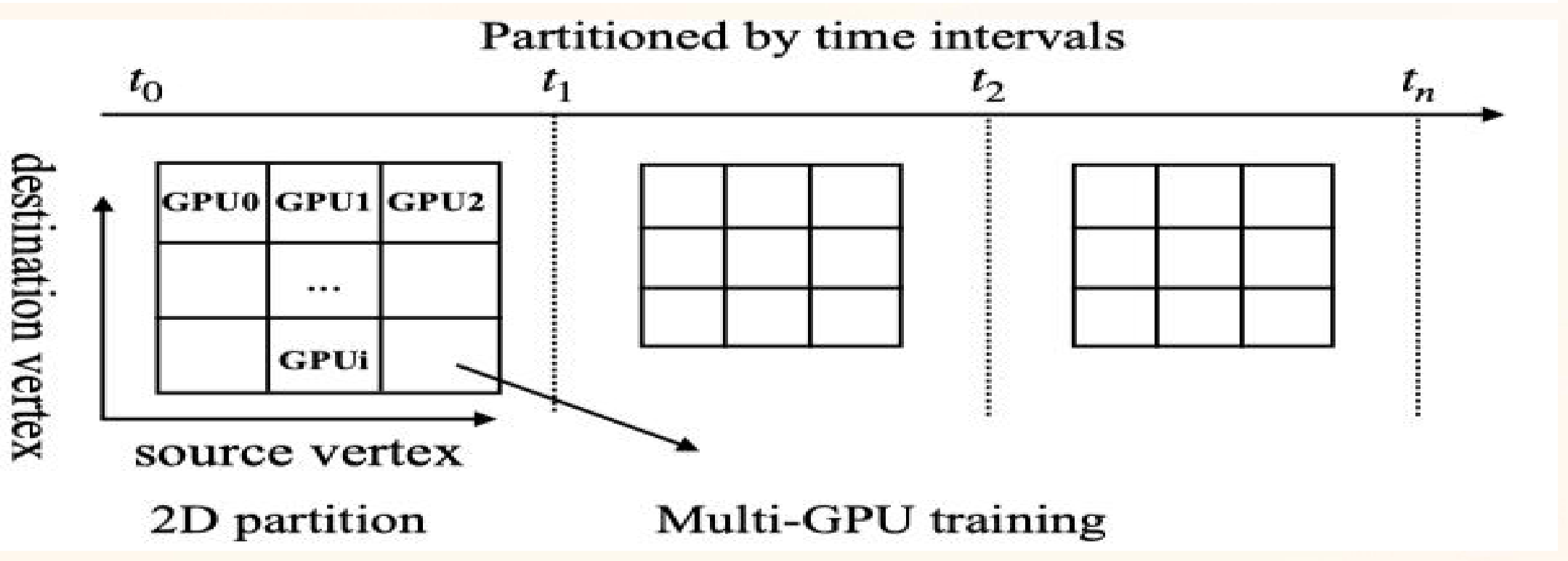


(a)WS-VR 统一虚拟地址通信(UVA)机制

(b)Groute 缓冲区环形链路通信机制

系统实现：面向异构互联架构的最优数据链路





Multi-GPU数据混合通信框架

- ❖内部链路感知（链路拓扑、连通和路由)系统优化
- ❖混合粒度数据链路：UVA（点）、NCCL（块）

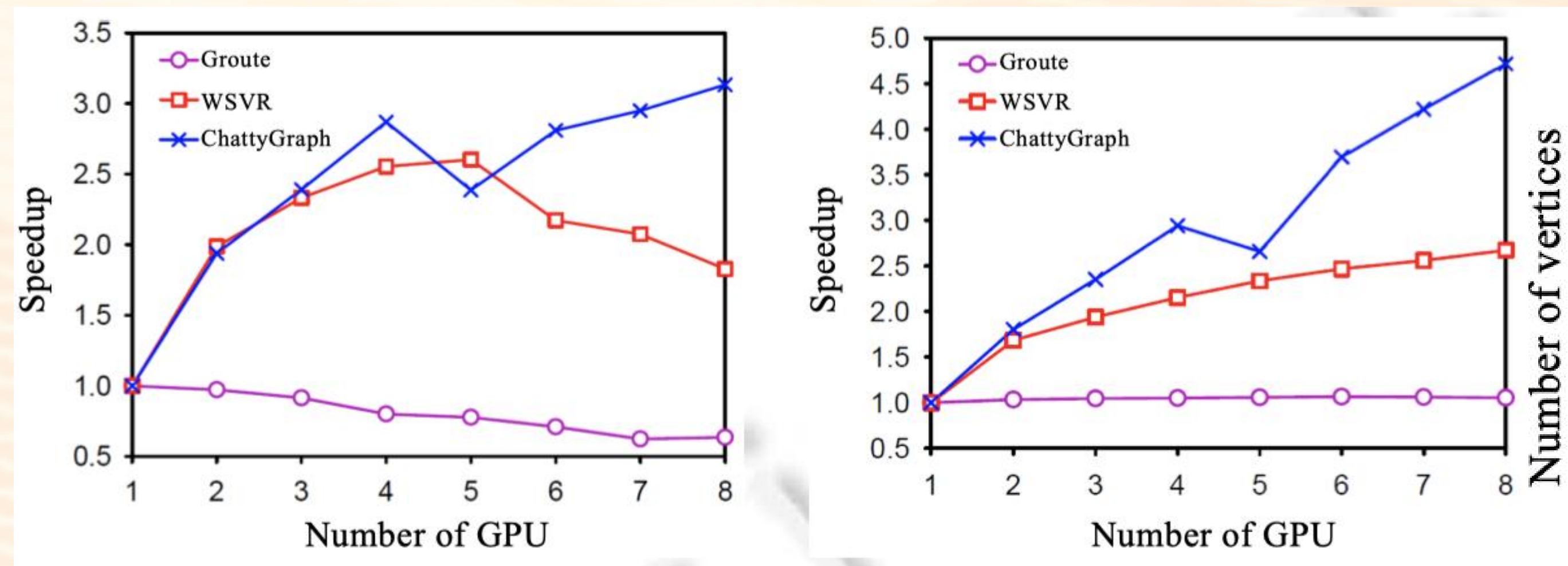
ChattyGraph：多协处理器图计算系统

- ❖编程框架：以边为中心的GAS接口
- ❖设备层级内存数据I/O：顺序化子图分块
- ❖子图均衡负载：边集均衡、负载窃取

T-GCN：异构协处理器GNN系统

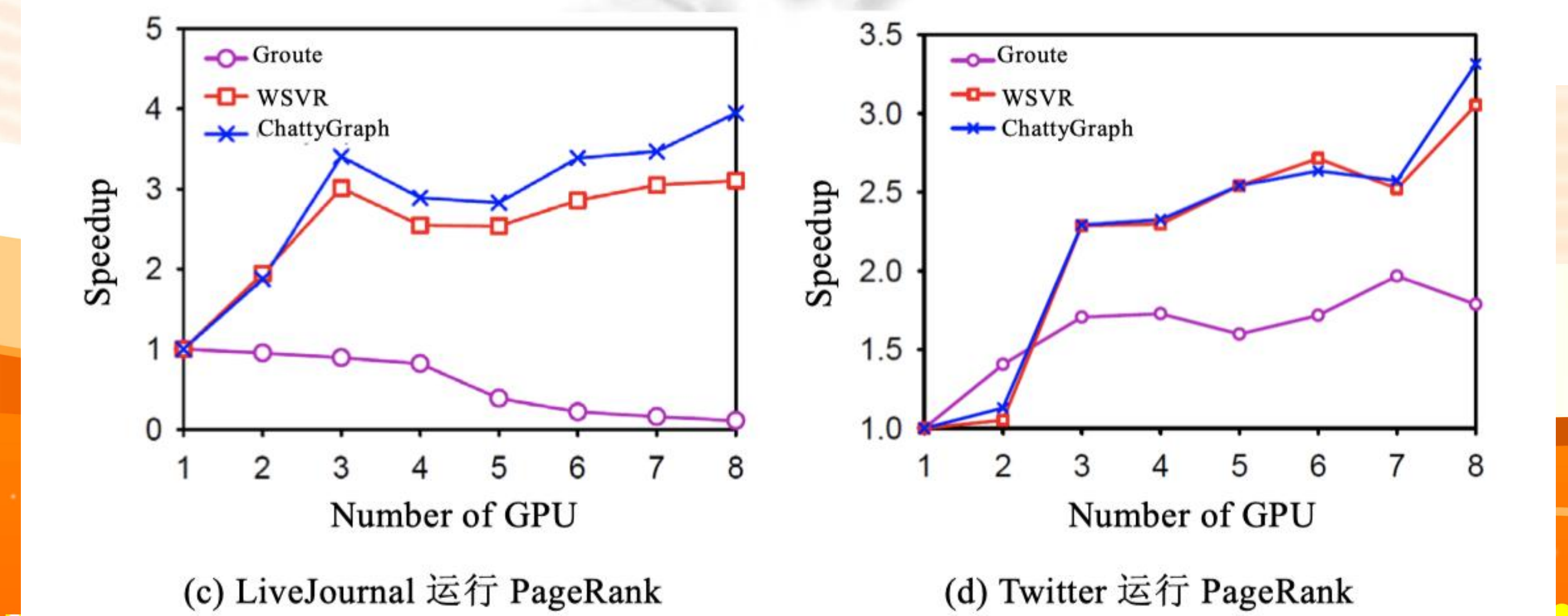
- ❖多GPU下图神经网络训练系统
- ❖设备Embedding本地缓存优化
- ❖NVLink链路最大化的分布训练任务调度

实验评估验证：Effectiveness&Micro-Benchmark



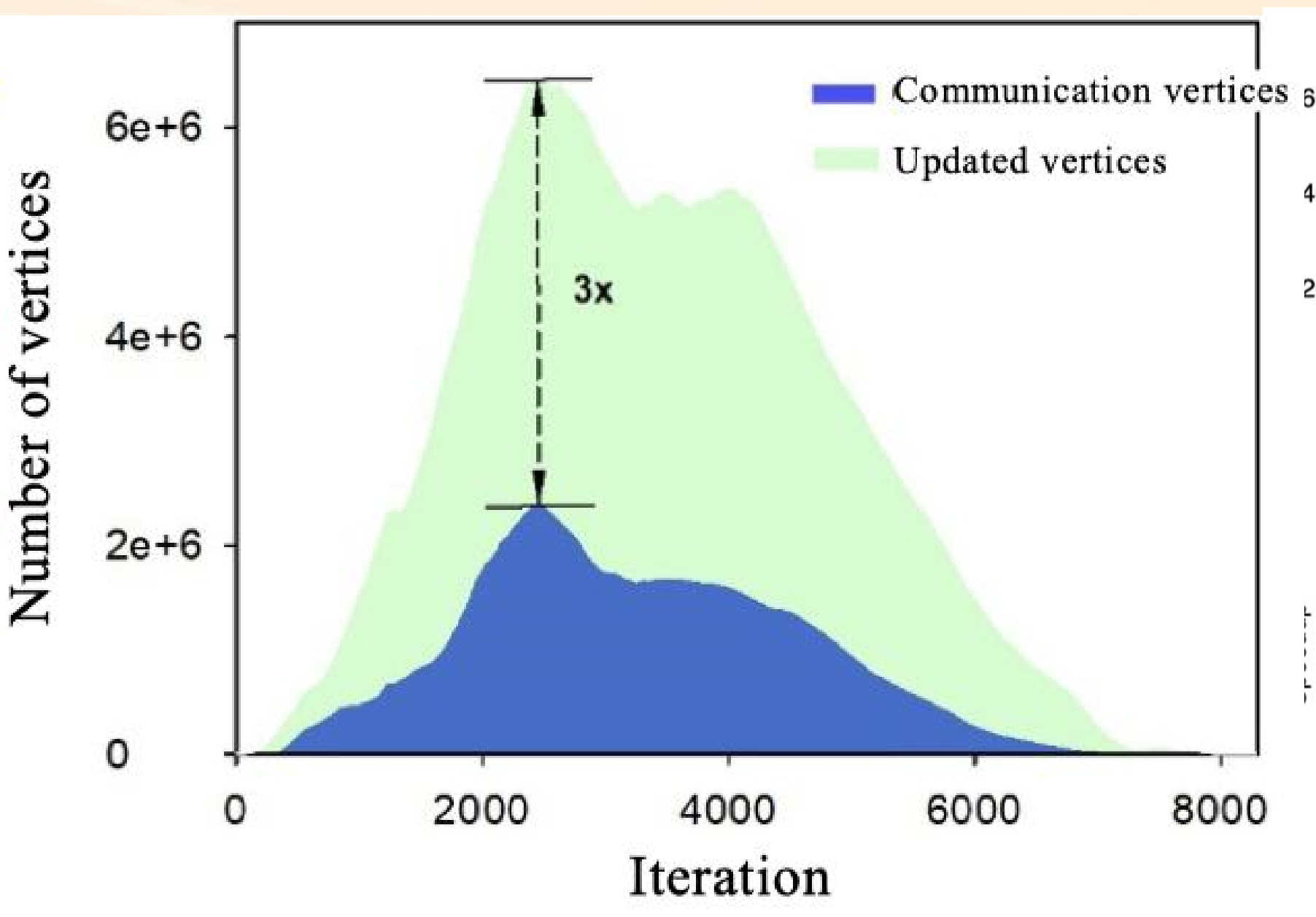
(a) 在 USA-road 运行 BFS

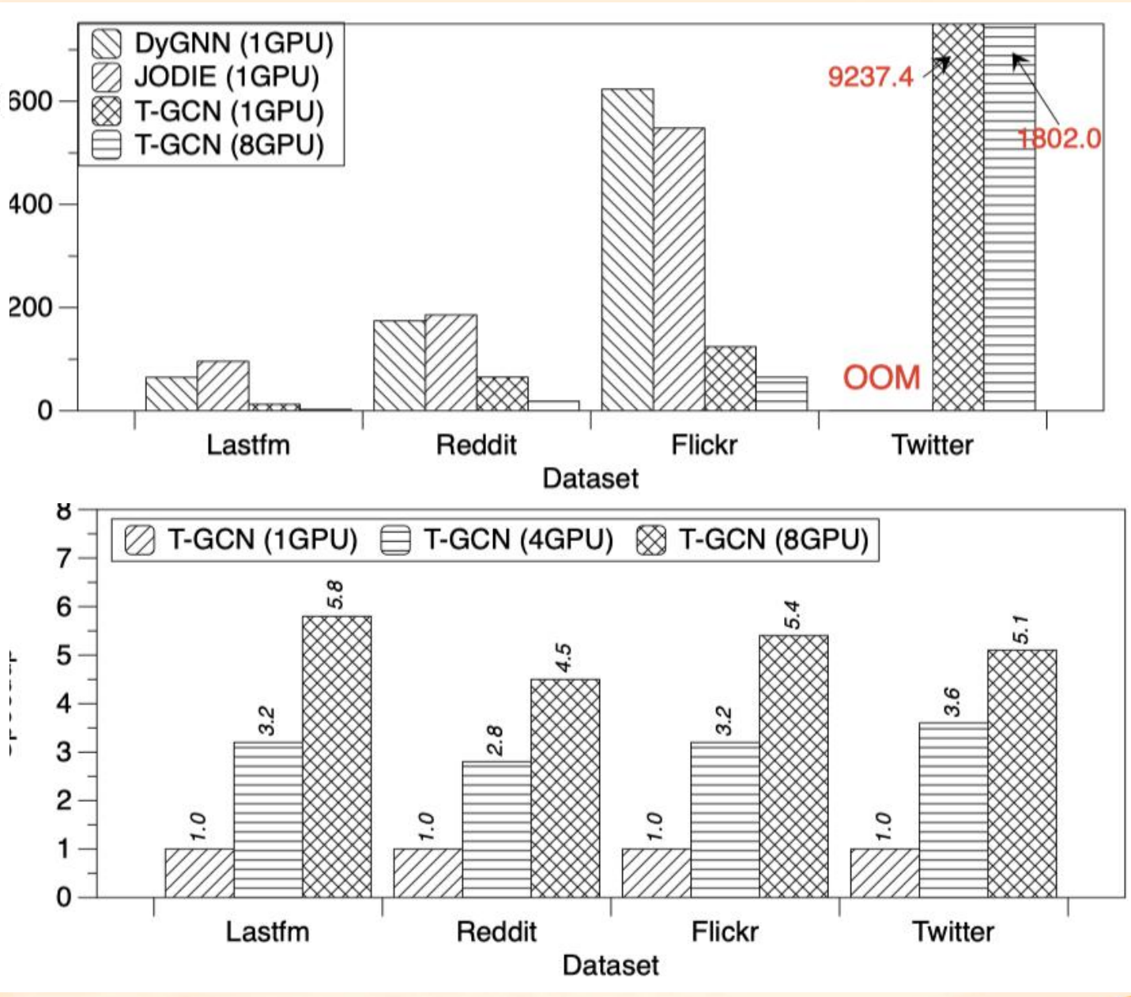
(b) 在 osm-eur 运行 SSSP



(c) LiveJournal 运行 PageRank

(d) Twitter 运行 PageRank





数据链路降低3X开销

系统性能加速比（2-3X）与可扩展性评估