

Bipartite matching under communication constraints

Moonmoon Mohanty, Gautham Bolar, Preetam Patil, Parimal Parag
Ayalvadi J. Ganesh, Jean-Francois Chamberland
Stochastic processes on networks workshop, Aug 03, 2026



CENTRE FOR
NETWORKED INTELLIGENCE
Indian Institute of Science

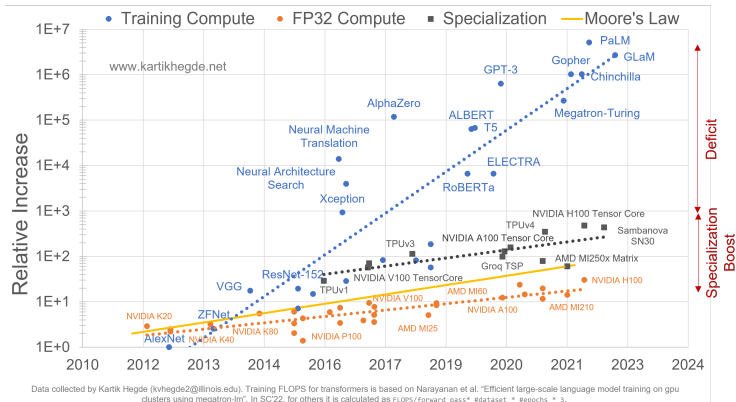


Qualcomm



The growing performance deficit

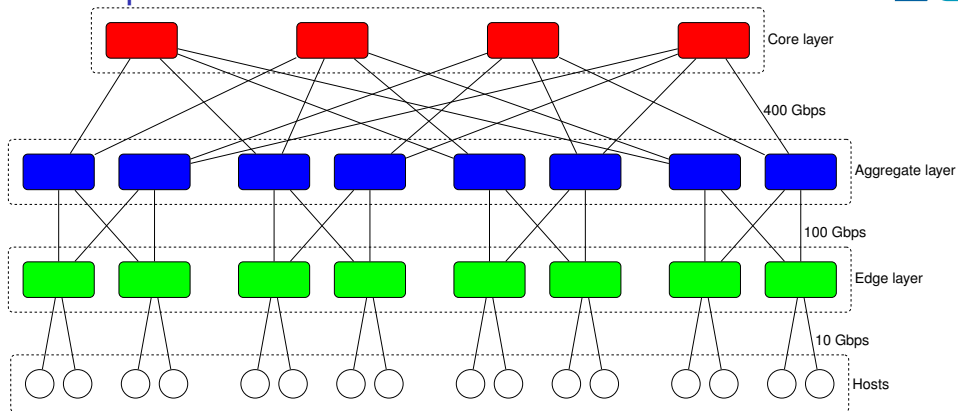
Deep learning compute demand¹



- ▶ End of Moore's law and Dennard scaling
- ▶ Can no longer keep adding more transistors and increasing core frequencies
- ▶ We're accustomed to increasing compute achievable only by scaling out through data center network (DCN)

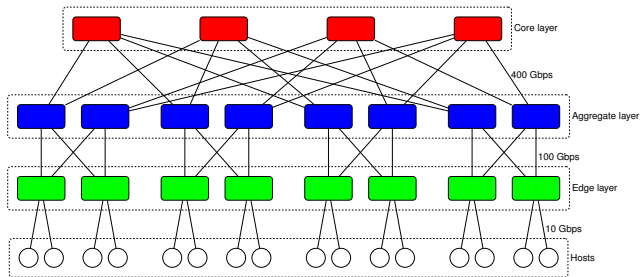
¹ Image credit: <https://kartikhegde.substack.com/p/accelerating-deep-learning-in-the>

DCN for compute



- ▶ DCN traffic comprises of
 - ▶ Latency sensitive compute applications (short messages)
 - ▶ Bandwidth hungry applications (long messages)
- ▶ The link from host to edge layer switch is the bottleneck
- ▶ End-hosts can be senders and receivers

Why not use TCP for networked compute?



Inadequacies of TCP for networked compute

- ▶ Not all messages are equal (TCP treats them equally)
- ▶ TCP congestion control operates at RTT (scale of microseconds in DCN)
 - ▶ Too high a latency for networked compute
 - ▶ High DCN bandwidth leads to large queue buildup and latency

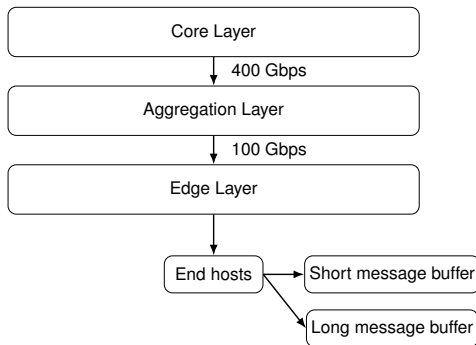
- ▶ Scheduling algorithms for input-queued switches in high-speed LANs and WAN routers
 - ▶ Parallel iterative matching (PIM)²
 - ▶ Iterative round robin scheduling algorithm (iSLIP),³ an improvement over PIM
- ▶ Transport protocols for networked compute
 - ▶ Priority based receiver driven Homa⁴ improvement over pHost, NDP
 - ▶ Data center parallel iterative matching (dcPIM)⁵

²Thomas E Anderson, Susan S Owicki, James B Saxe, and Charles P Thacker. High-speed switch scheduling for local-area networks. *ACM Transactions on Computer Systems*, 11(4):319–352, Nov. 1993.

³Nick McKeown. The iSLIP Scheduling Algorithm for Input-queued Switches. *IEEE/ACM Transactions on Networking*, 7(2):188–201, Apr. 1999.

⁴Behnam Montazeri, Yilong Li, Mohammad Alizadeh, and John Ousterhout. 2018. Homa: A Receiver-driven Low-latency Transport Protocol Using Network Priorities. *ACM SIGCOMM 2018*.

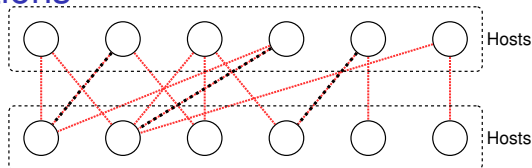
⁵Qizhe Cai, Mina Tahmasbi Arashloo, and Rachit Agarwal. dcPIM: Near-optimal proactive datacenter transport. *ACM SIGCOMM 2022*.



Proposals for transport protocols for compute

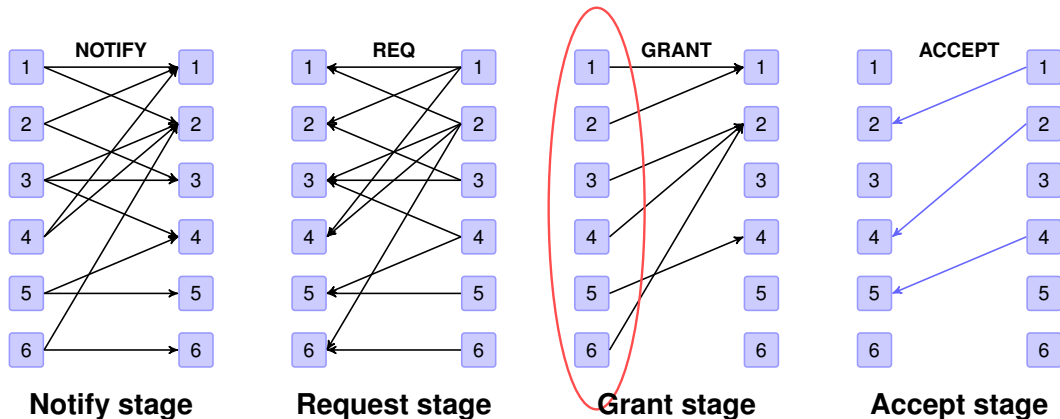
- ▶ Distinguish among control, short compute messages, and long persistent messages
- ▶ Short messages are prioritized and no congestion control for them
- ▶ Congestion control for long messages via receiver grant mechanism
- ▶ Matching of sender receiver pairs for efficient utilization of bottleneck edge links

Modeling assumptions



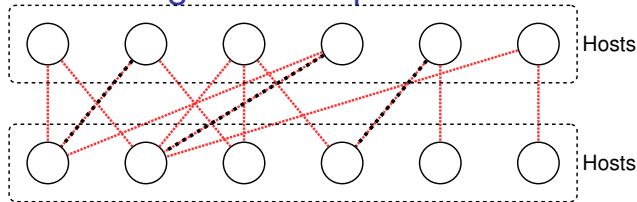
- ▶ Time is divided into discrete slots in order of microseconds
- ▶ Each slot consists of a matching and data transmission phase
- ▶ Matching phase
 - ▶ forming intention graph based on senders with long messages for receivers
 - ▶ collecting neighborhood information
 - ▶ matching algorithm
- ▶ Data transmission phase: parallel long message transmission for each matched pair
- ▶ Intention graph is modeled as a D -out random graph
 - ▶ $D = \text{Bin}(N, d/N)$ yields Erdős-Rényi graph
 - ▶ $D = d$ yields deterministic out degree
- ▶ **Goal: Establish a large bipartite matching in each time slot**

Control message exchange during matching phase



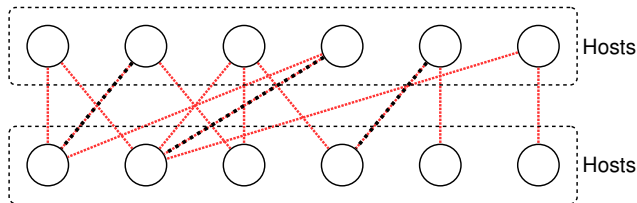
- ▶ Bipartite matching sender receiver pairs is at the heart of the problem
- ▶ Each round consists of matching and data transmission phases
- ▶ Distributed matching with local information to keep the matching phase small

Matching algorithms using Grant requests



Related works

- ▶ Centralized bipartite matching
 - ▶ Hopcroft-Karp: computes maximum matchings in polynomial time
 - ▶ Hungarian: polynomial-time algorithm widely used in applications like client selection or resource allocation to maximize total weight
- ▶ Distributed matching
 - ▶ Random bipartite matching
 - ▶ Uses one neighborhood information
 - ▶ Picks one of its intended receivers uniformly at random
 - ▶ Proposal: Degree-Biased (DB) (α) random bipartite matching
 - ▶ Uses two neighborhood information
 - ▶ Biased selection of one of its receivers

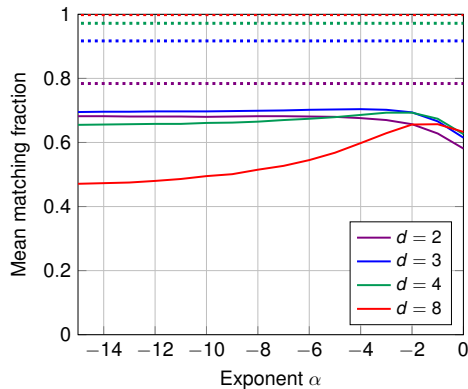


DB(α) algorithm

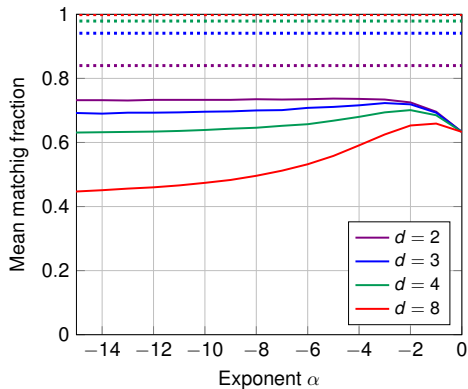
- ▶ Edge $u \sim v$ selection probability: $\deg_v^\alpha / \sum_{w: u \sim w} \deg_w^\alpha$
- ▶ $\alpha = 0$: uniform selection
- ▶ $\alpha < 0$: prefer low degree neighbors
- ▶ $\alpha = -\infty$: greedy selection

Mean matching fraction

Impact of changing exponent α



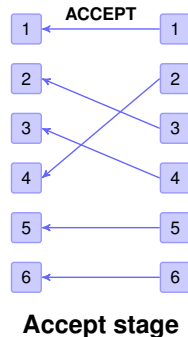
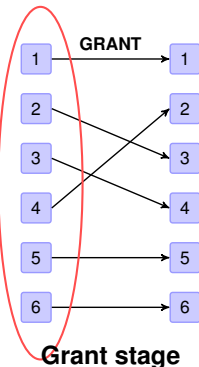
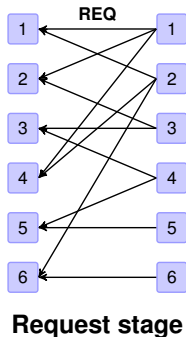
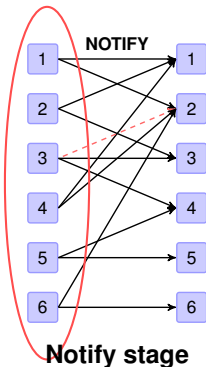
(a) $D \sim \text{Bin}(N, d/N)$



(b) Deterministic $D = d$

- Higher mean sender degree can counterintuitively reduce matching performance

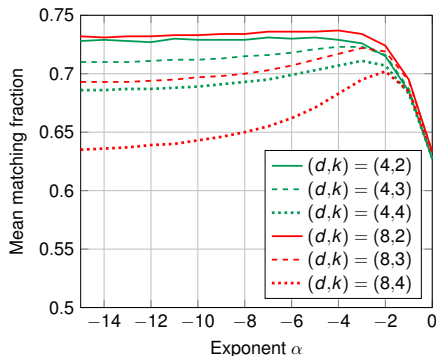
2CGS algorithm for notification and grant selection



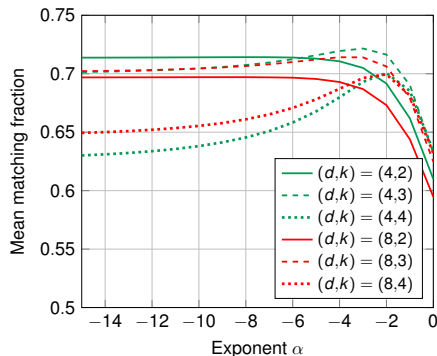
- ▶ Random thinning of one neighborhood in the notify stage
- ▶ Uses two neighborhood information built during the request stage
- ▶ Greedy selection of a minimum degree neighbor uniformly at random in grant stage

Mean matching fraction

Impact of thinning

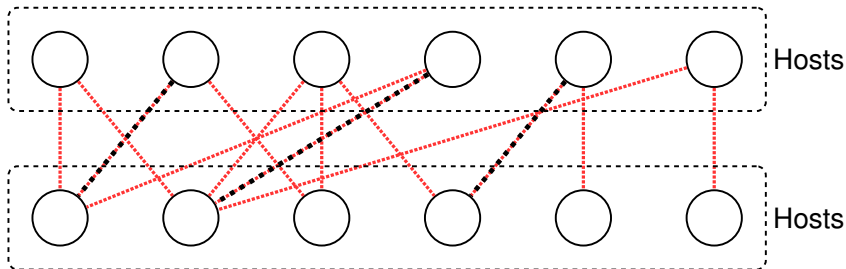


(a) $D \sim \text{Bin}(N, d/N)$ with $\max(k)$



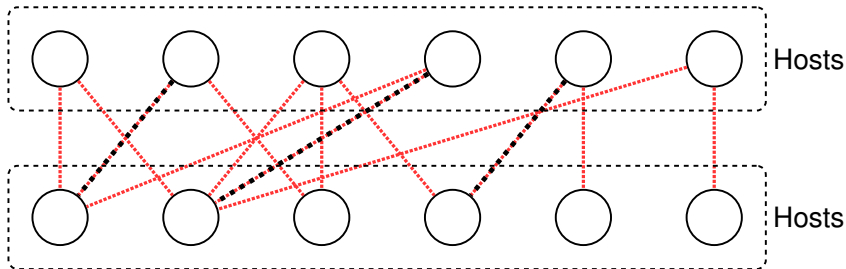
(b) Deterministic $D = d$ with $\text{Bern}(k/d)$

- ▶ Thinning to low out-degree improves decentralized matching performance
- ▶ $\max(2)$ yields the strongest results



Uniform selection

- ▶ Probability of a receiver being picked by any sender is $P\{D > 0\} / N$
- ▶ Mean matching fraction $1 - \left(1 - \frac{P\{D > 0\}}{N}\right)^N \approx 1 - e^{-P\{D > 0\}}$



Degree distribution for receivers

- ▶ Probability for an edge to be present $\mathbb{E}D/N$
- ▶ Receiver degree $\text{Bin}(N, \mathbb{E}D/N)$ irrespective of distribution D

Size biased distribution

- ▶ Conditional probability of sender degree being k given connected: $kP\{D = k\} / \mathbb{E}D$.

Greedy selection

- ▶ Under some technical conditions, neighbors of a receiver have independent size biased degree distributions
- ▶ Mean matching fraction is lower bounded by

$$1 - \sum_{s \geq 0} e^{-\mathbb{E}D} \frac{(\mathbb{E}D)^s}{s!} (1 - f(s))^s$$

where π and Π are pmf and cdf for Poisson random variables with parameter $\mathbb{E}D$ and

$$f(s) \triangleq \frac{G_D(\bar{\Pi}_{s-2}) - G_D(\bar{\Pi}_{s-1})}{\mathbb{E}D \pi_{s-1}}.$$

Feature of the proposed distributed 2CGS matching algorithm

- ▶ Based on two neighborhood information
- ▶ Theoretical analysis for greedy selection and D -out random graphs in both sparse and dense regimes
- ▶ Improves the mean matching size and hence throughput for long messages
- ▶ Reduces control message overheads, improving throughput for short messages as well

Reference

M. Mohanty, G. Bolar, P. Patil, A. J. Ganesh, J. F. Chamberland, P. Parag, “Bipartite matching under communication constraints”, *arXiv:2604.10744*, Jun 2026.