Miniature: Fast AI Supercomputer Networks Simulation on FPGAs

Yicheng Qian[1,2], **Ran Shu**[2], Rui Ma[2], Yang Wang[2], Derek Chiou[2], Nadeen Gebara[2], Luca Piccolboni[2], Miriam Leeser[1] and Yongqiang Xiong[2]

[1]Northeastern University, [2]Microsoft Research, [3]UT Austin, [4]Microsoft

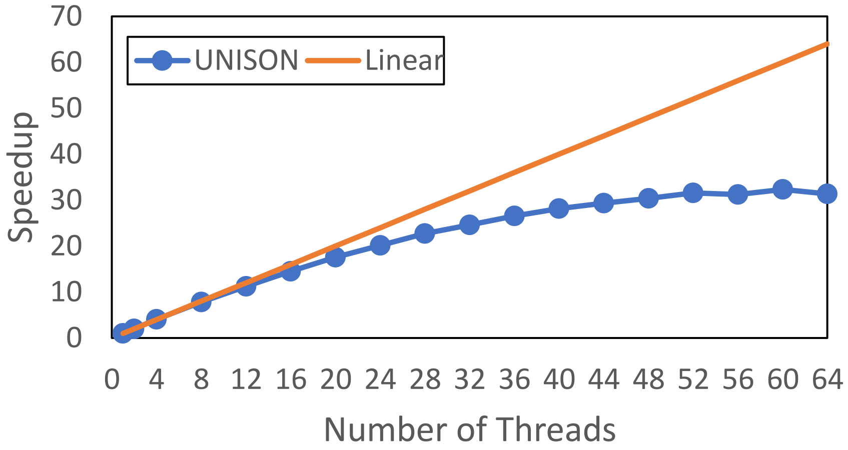
2025-08-08

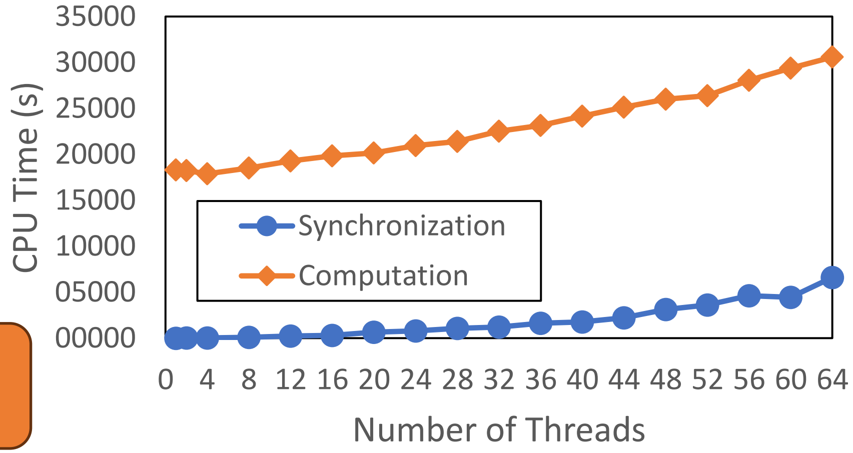
Requirements of DES

- Costly to run real testbed
  - Searching parameter space
  - Planning new clusters
  - Improving infrastructure e.g., collective communication and transport
- Simulation is the cost effective way
- **Need high fidelity discrete event simulation (DES) for network**
  - Improve DES efficiency is important
  - DONS[SIGCOMM'23], UNISON[EuroSys'24], SimAI[NSDI'25]

Performance Issues of DES

- Poor single thread performance
  - Inefficient general computation
  - **2,760,000 times slower**
- Cannot scale linearly to multiple threads
  - **Up to 32x speedup with 64 threads**
  - **85,000 times slower**
  - Due to increased synchronization overhead and cache contentions





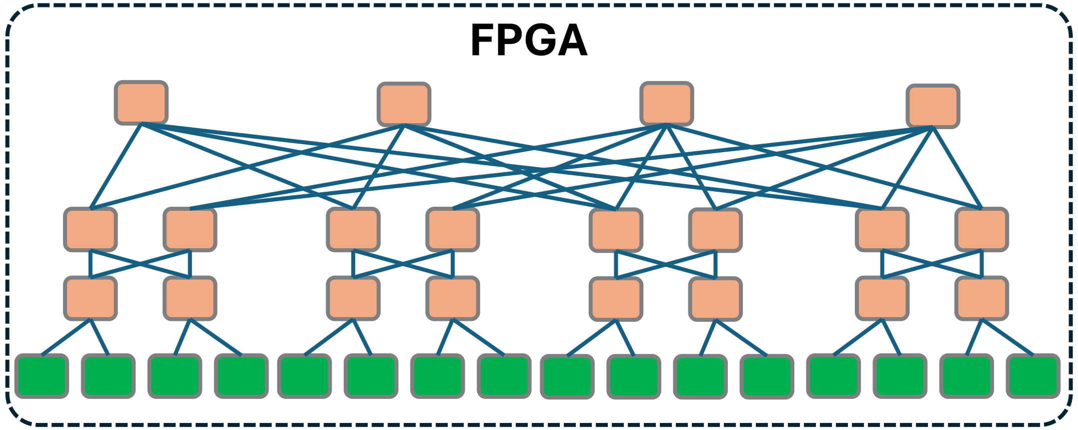
Impossible for large scale in-terms of time and money

Simulation settings: UNISON [1]. 1024-node Fat-tree topology with 400 Gbps link. Tree-based all-reduce operation.

[1] Bai, Songyuan, et al. "Unison: A Parallel-Efficient and User-Transparent Network Simulation Kernel." EuroSys 2024

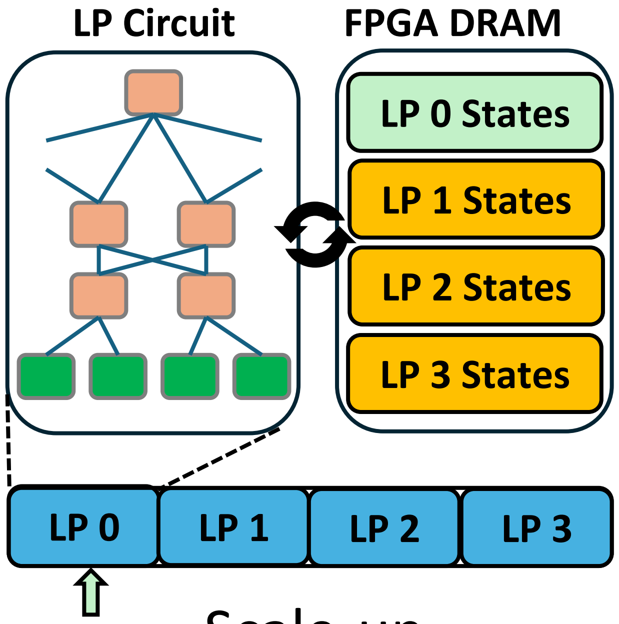
From Software-based to Miniature

- Software-based DES limitations
  - Inefficient network protocol processing
  - High event coordination overhead due to virtual timer
  - Inefficient to scale
- A possible miniature world on FPGA
  - Specialized circuit
  - Events happens in (near) real time to minimize event overhead
  - Naturally parallel

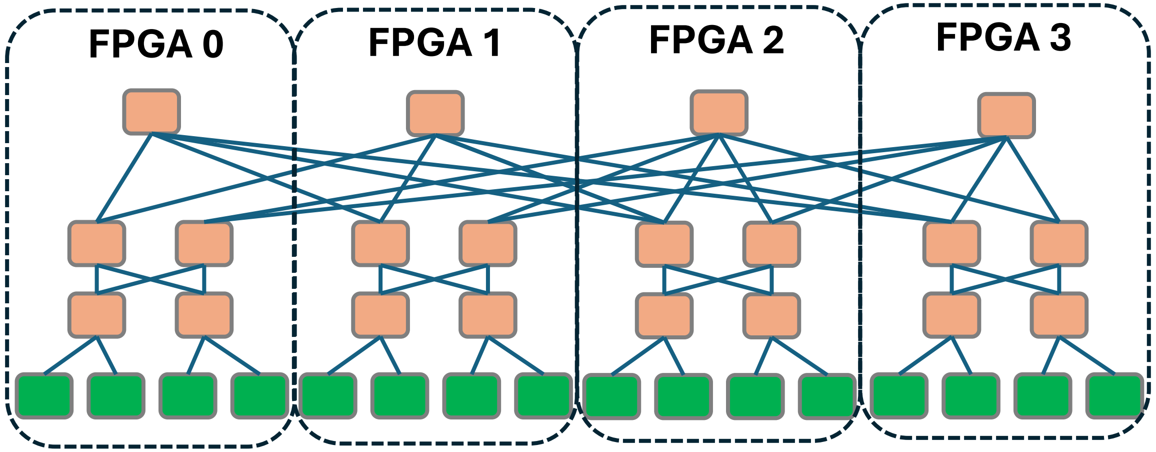


Scalable Simulation

- **Can only fit ~1000 nodes into an FPGA**
- Solution: **multiplexing & scale-out**
- Multiplexing by context switch
  - Only switch data but not computation
  - Minimal link delay as the switch time
  - Tired multiplexing using BRAM and DRAM
- Scale-out through inter-FPGA communication
  - Hiding Inter-FPGA delay naturally



Scale-up



Scale-out

Multiplexing vs. Scale-out

Multiplexing Cons

- DRAM bandwidth bottleneck
  - Use HBM can greatly mitigate the issue but still costly
- Higher simulation time

Scale-out Cons

- Inter-FPGA bandwidth bottleneck
  - Even if only transmit minimal packet header fields
- Requires large number of FPGAs

Friends not foes: should use as much as FPGAs to get optimal performance

Preliminary Results – FPGA Resource Usage

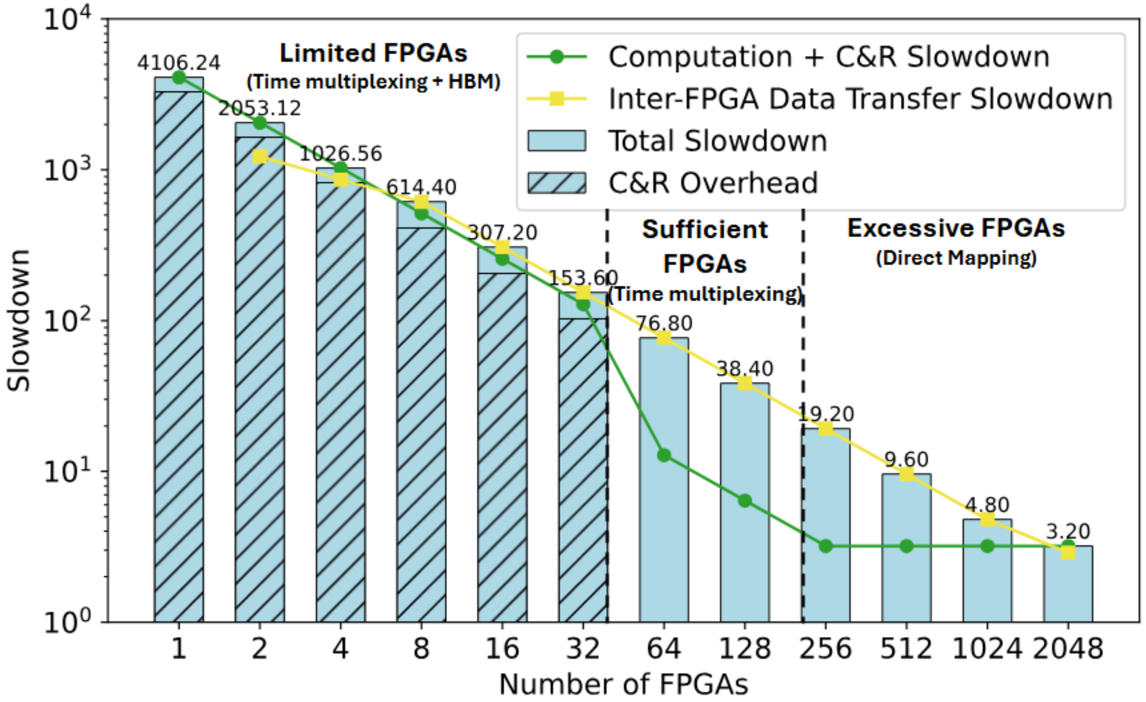
	Module	Logic Cells	BRAM	URAM
Endpoint	Basic EP	213 (0.054%)	6 (0.3%)	0
	PFC	17 (0.004%)	0	0
Switch	64-port Switch	36172 (2.770%)	0	6 (0.625%)
	PFC	147 (0.034%)	0	0

- FPGA: AMD Xilinx Alveo U280

Extremely efficient for simulating network components

~500 EPs and 20 switches on a single FPGA

Preliminary Results – Estimated Performance



Bottleneck:

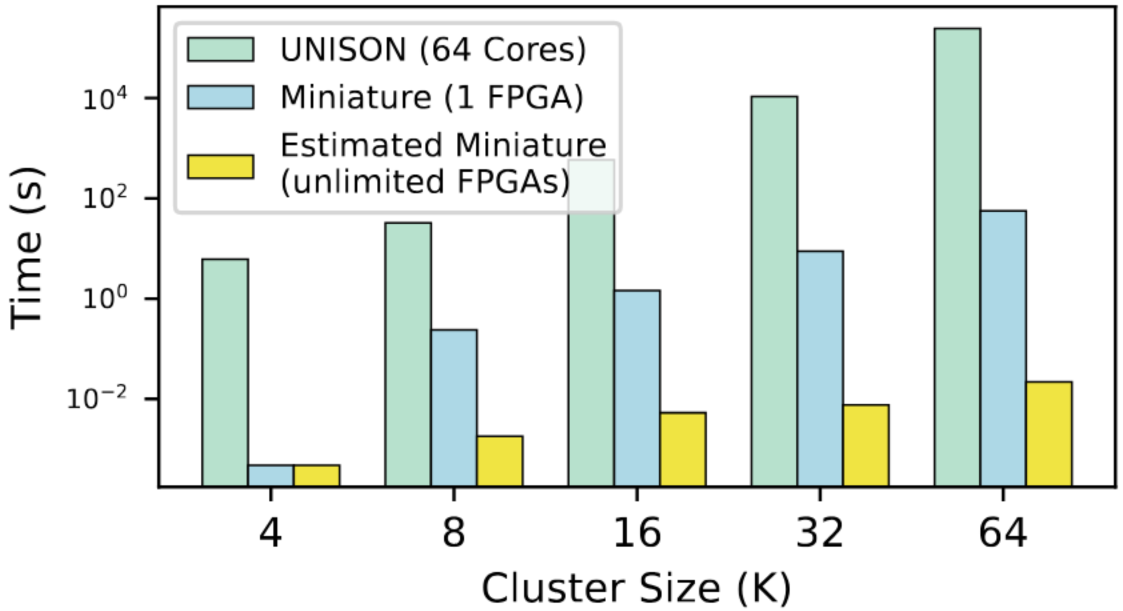
Multiplexing for 1-4 FPGAs

Inter-FPGA bandwidth for 8-1024 FPGAs

Computation for 2048 FPGAs

Settings: 65,536-node Fat-tree topology with 400 Gbps link.

Preliminary Results – Performance Comparison



Miniature can **speed-up** simulation up to **4,332 x** with single FPGA, and **5.6M x** with 2,048 FPGAs

Conclusion

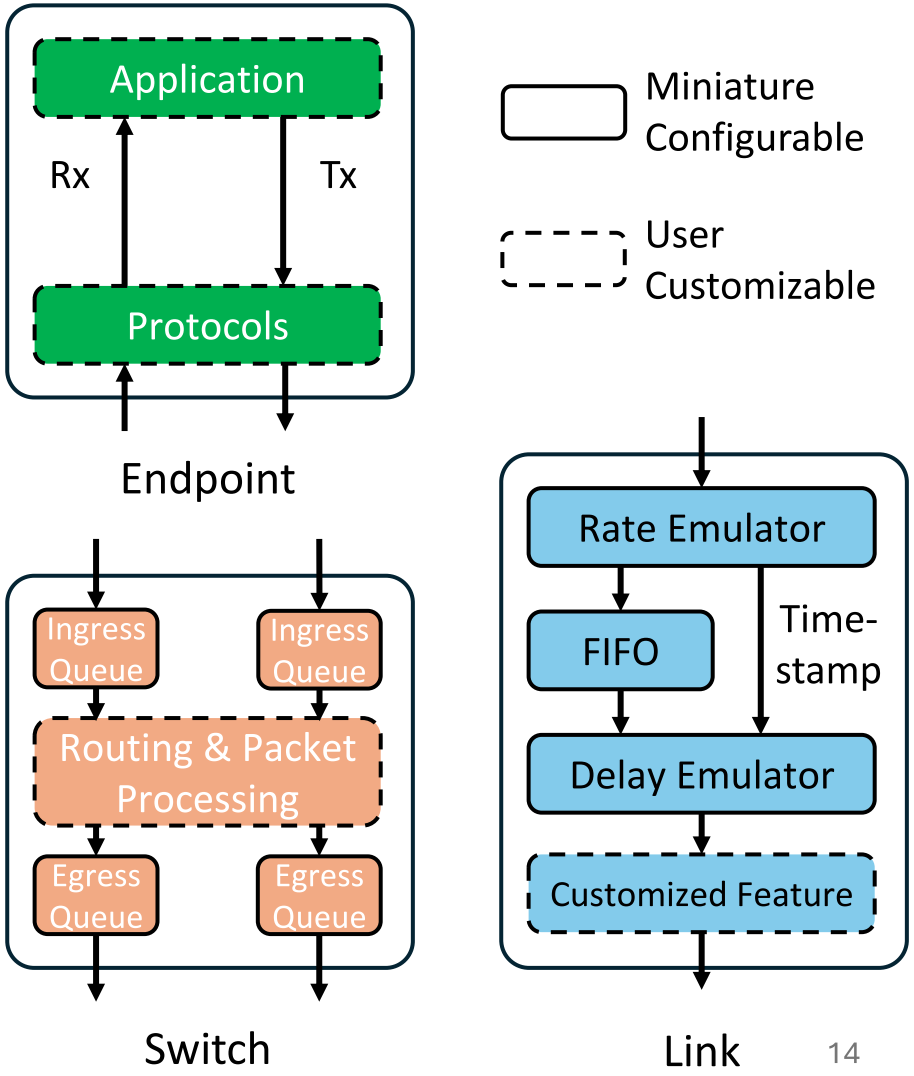
- Software-based DES has a speed upper bound
  - Slow single thread performance
  - Cannot scale linearly
- Miniature: FPGA-based AI network simulation
  - Scaling through multiplexing and scale-out
  - Over 4,000x speed-up compared with UNISON using single FPGA

Thank you!

Backup

Simple and Efficient Network Abstraction

- Highly efficient
  - Only include basic requirements for network simulation
  - Special design for FPGA characteristics
- Highly configurable and customizable
  - Efficient configurable logic
    - E.g., link bandwidth and delay
  - User customizable logic
    - E.g., customized routing or congestion control



Application
Rx
Tx
Protocols
Endpoint
Miniature Configurable
User Customizable
Ingress Queue
Ingress Queue
Routing & Packet Processing
Egress Queue
Egress Queue
Switch
Rate Emulator
FIFO
Time-stamp
Delay Emulator
Customized Feature
Link