SimBricks: End-to-End Network System Evaluation with Modular Simulation

Hejing Li

Jialin Li

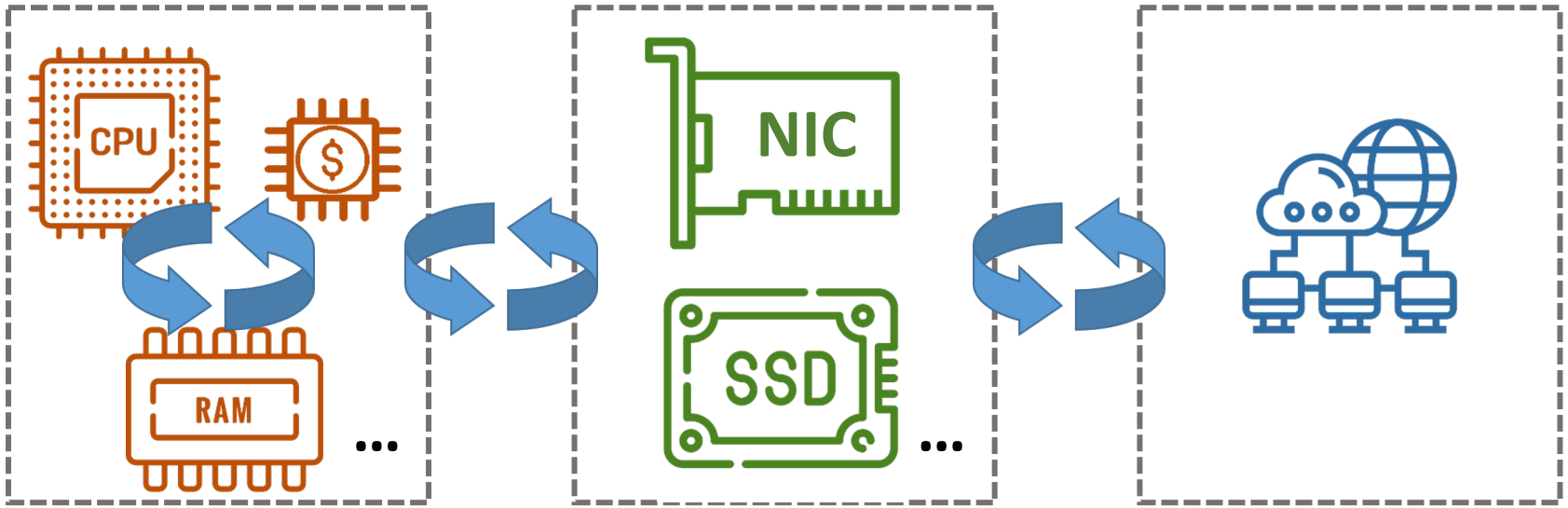
Antoine Kaufmann

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

NUS
National University
of Singapore

End-to-End Measurements Are Essential

- The metrics we care about are full system “end-to-end” properties
  - Throughput, latency…
- Many factors in system component affect the overall behavior
  - Host architecture and hardware device performance etc.



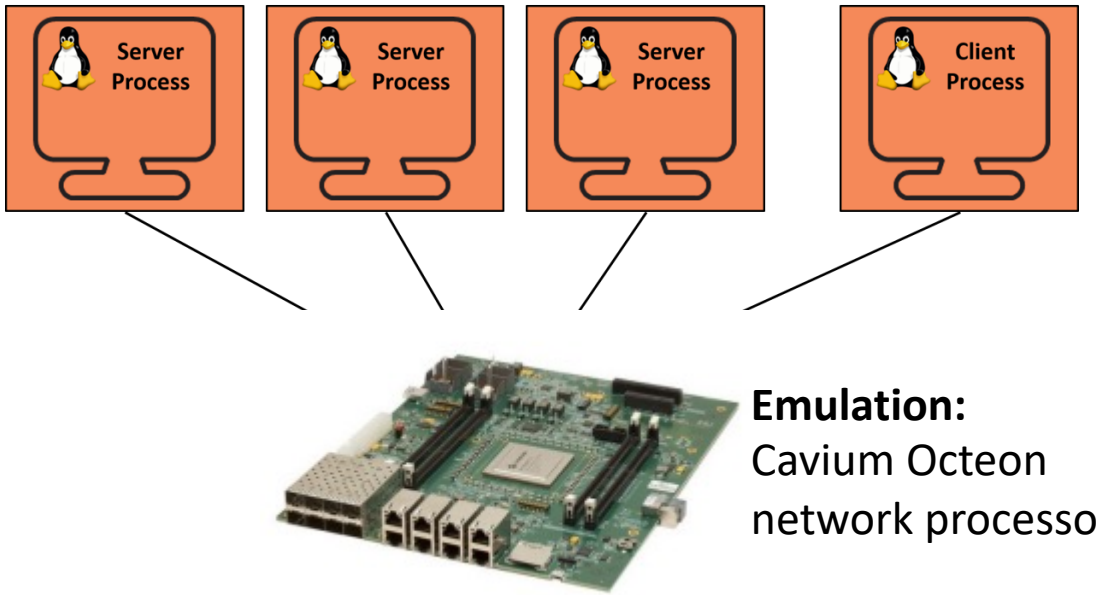
E-to-E Measurements Are Often Not Possible

- Commercial h/w is not available at publication time
  - E.g. Programmable switches

Just Say NO to Paxos Overhead:
Replacing Consensus with Network Ordering

Jialin Li Ellis Michael Naveen Kr. Sharma Adriana Szekeres Dan R. K. Ports
University of Washington

[OSDI'16]



Emulation:
Cavium Octeon
network processor

E-to-E Measurements Are Often Not Feasible

- Commercial h/w is not available at publication time
  - E.g. Programmable switches
- Propose new ASICs or h/w extensions


A Cloud-Scale Acceleration Architecture

Adrian M. Caulfield Eric S. Chung Andrew Putnam
Hari Angepat Jeremy Fowers Michael Haselman Stephen Heil Matt Humphrey
Puneet Kaur Joo-Young Kim Daniel Lo Todd Massengill Kalin Ovtcharov
Michael Papamichael Lisa Woods Sitaram Lanka Derek Chiou Doug Burger

Microsoft Corporation

[MICRO'16]

Require large scale networks
(e.g. 10s - 100s hosts)


Approximating Fair Queueing on Reconfigurable Switches

*Naveen Kr. Sharma** *Ming Liu** *Kishore Atreya*† *Arvind Krishnamurthy**

[NSDI'18]

Simulation to the Rescue?

There are well established simulators for…

Host

Hardware

Network

None of them covers full end-to-end system

Simulation to the Rescue?

There are well established simulators for…

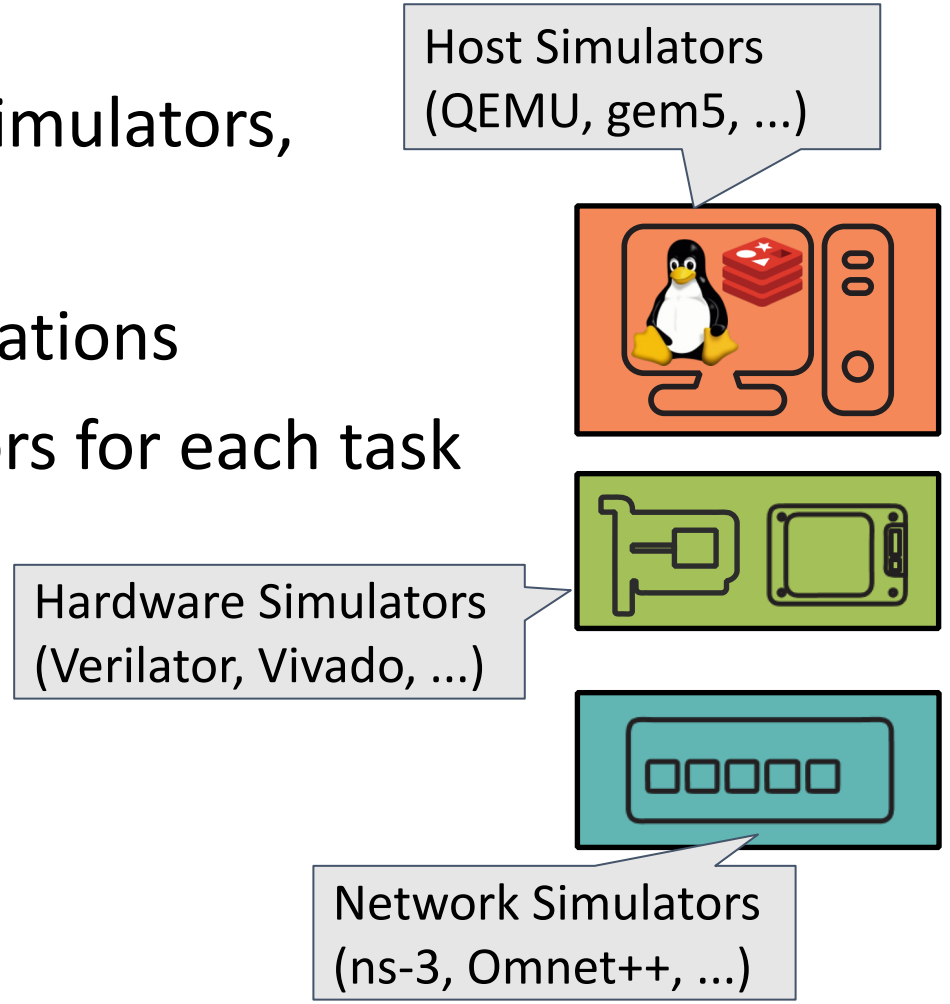
Host | Hardware | Network

Can we combine them into a full system simulation?

SimBricks: Modular E-to-E Network System Simulations

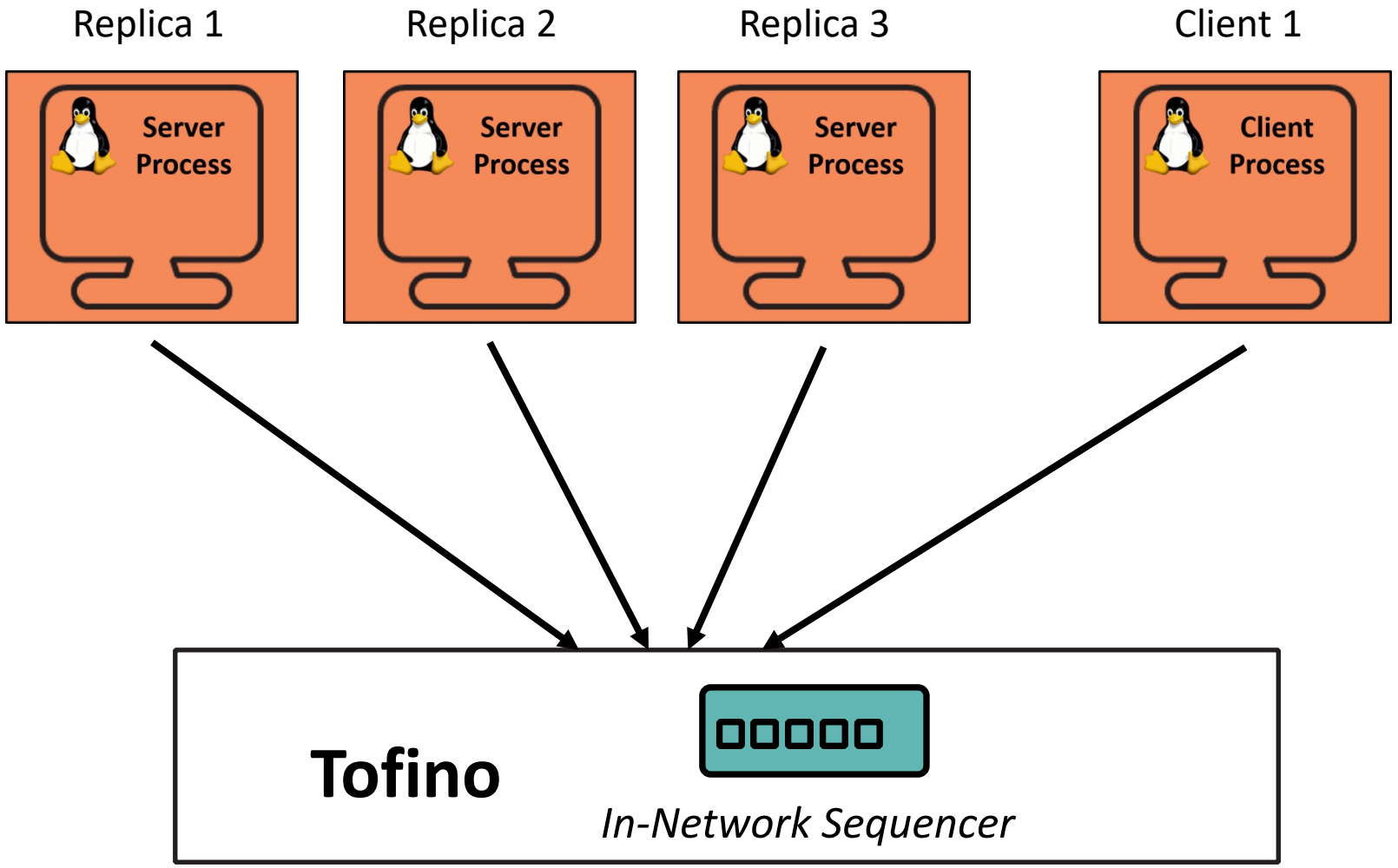
By combining host, device, and network simulators, we aim to

- Run complete HW and SW implementations
- Pick the right combination of simulators for each task
- Scale to large systems
- Enable easy simulator integration



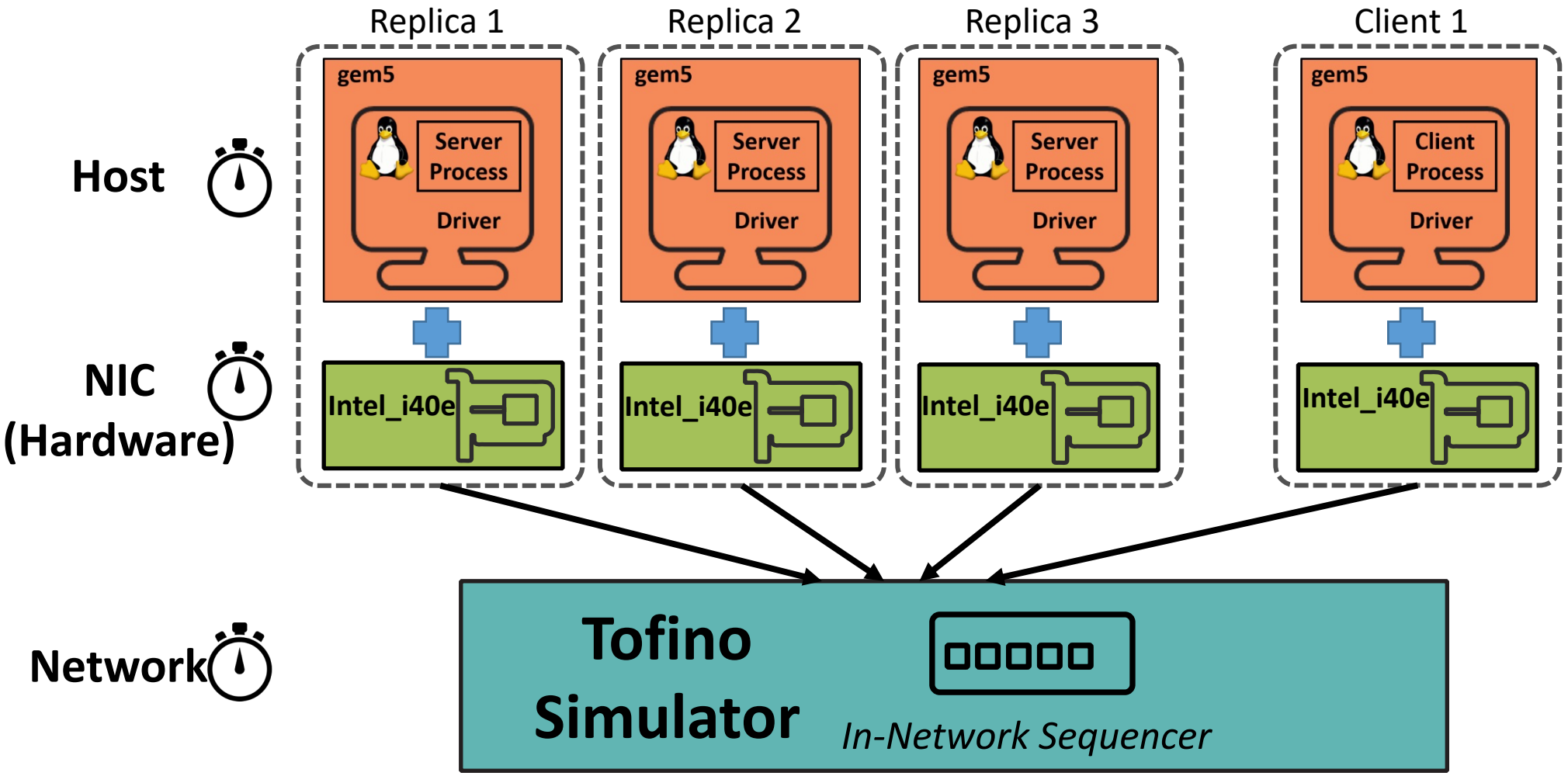
Composing E-to-E Simulation with SimBricks

- NOPaxos physical testbed configuration



Composing E-to-E Simulation with SimBricks

- NOPaxos SimBricks configuration



Technical Challenges

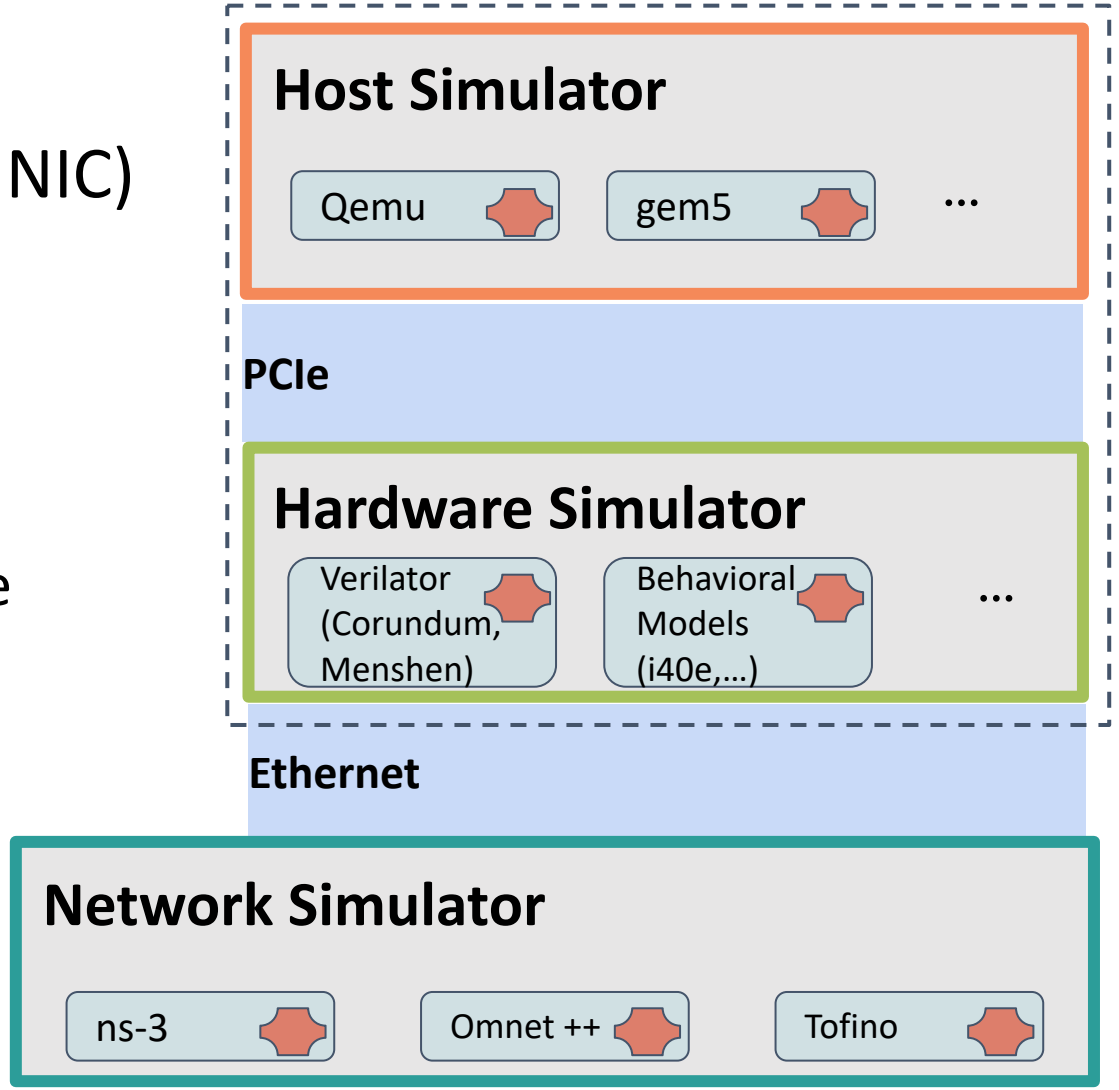
1. No interface for interconnecting with other simulators
2. Synchronization and communication overheads
3. Incompatible simulation models

Host Hardware Network

Fix Natural Component Simulator Interfaces

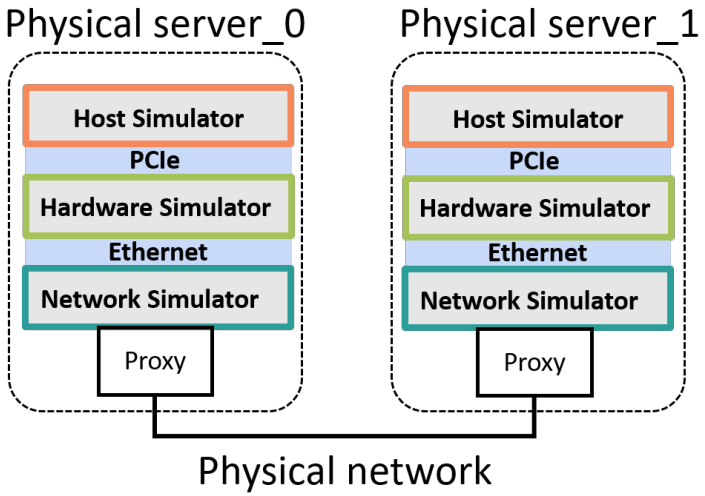
- ***PCIe***: Host - Device (PCI device. e.g. NIC)
- ***Ethernet***: NIC – Network

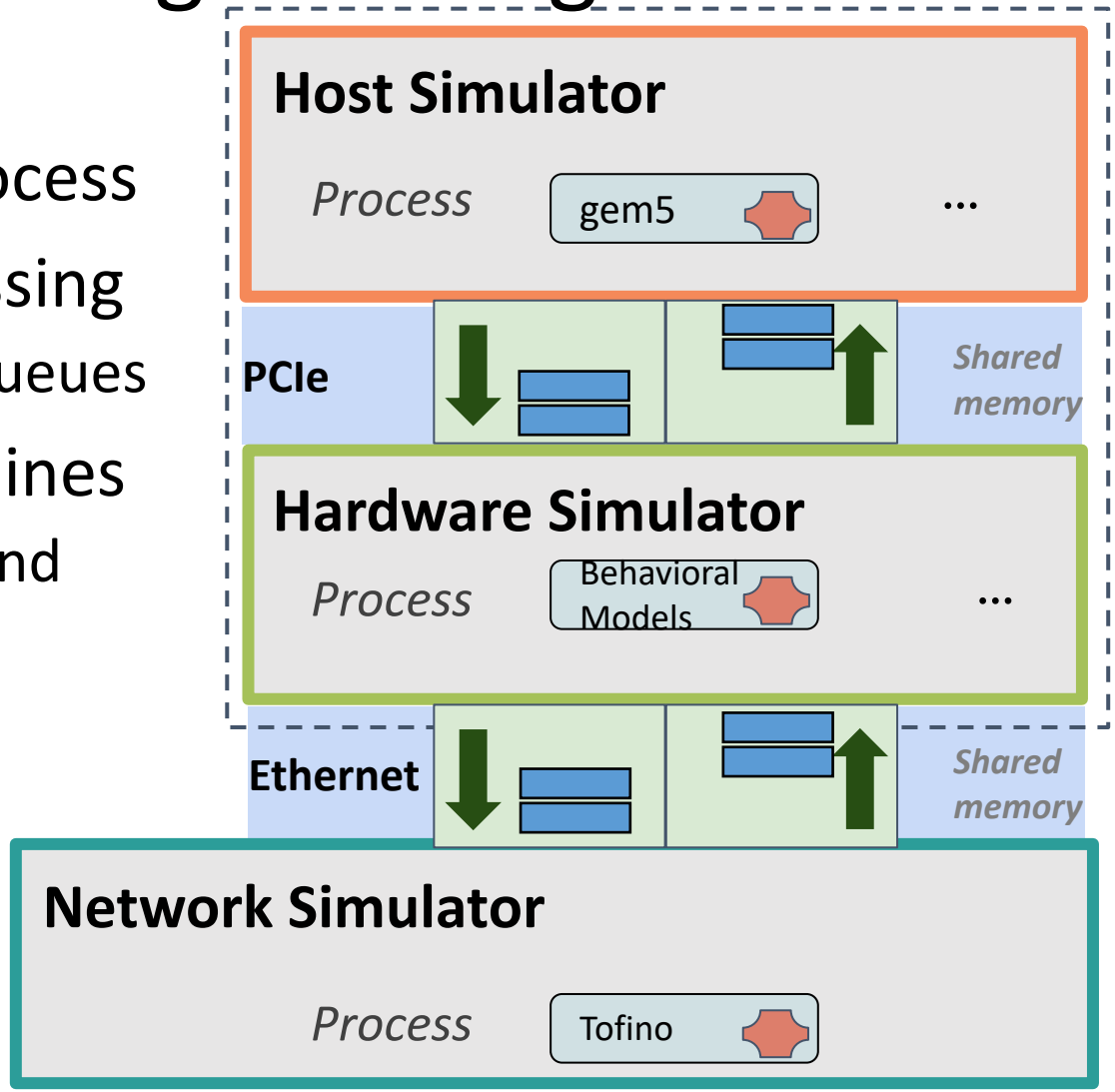
- ***SimBricks Adapter***:
  - Implements the components interface



Parallel Execution with Message Passing

- Run each simulator in individual process
- Communicate through message passing
  - Optimized pairwise shared memory queues
- Scaling out to several physical machines
  - Proxy: Translating between network and SHM queues





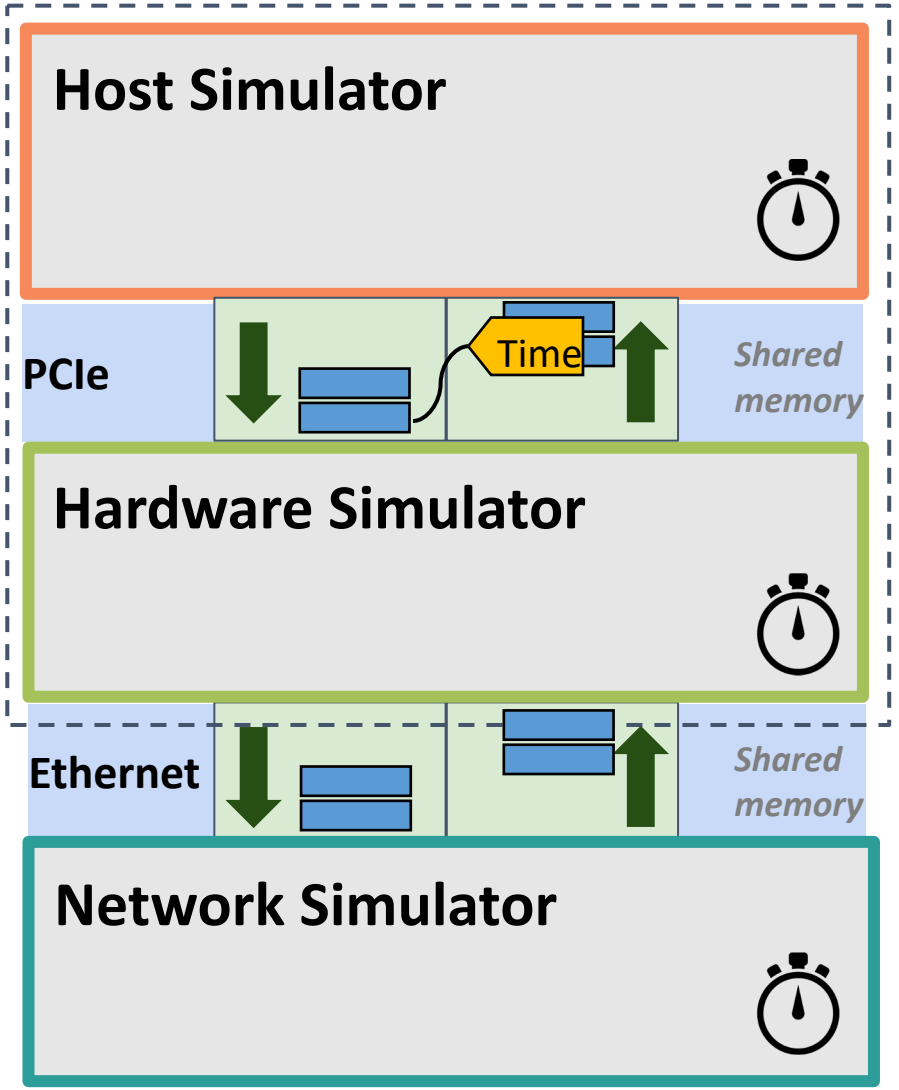
Simulator Synchronization Mechanism

Goals:

- Accurate simulation
  - Produce meaningful performance results
- Minimal synchronization overhead
  - Scalable synchronization

Observations:

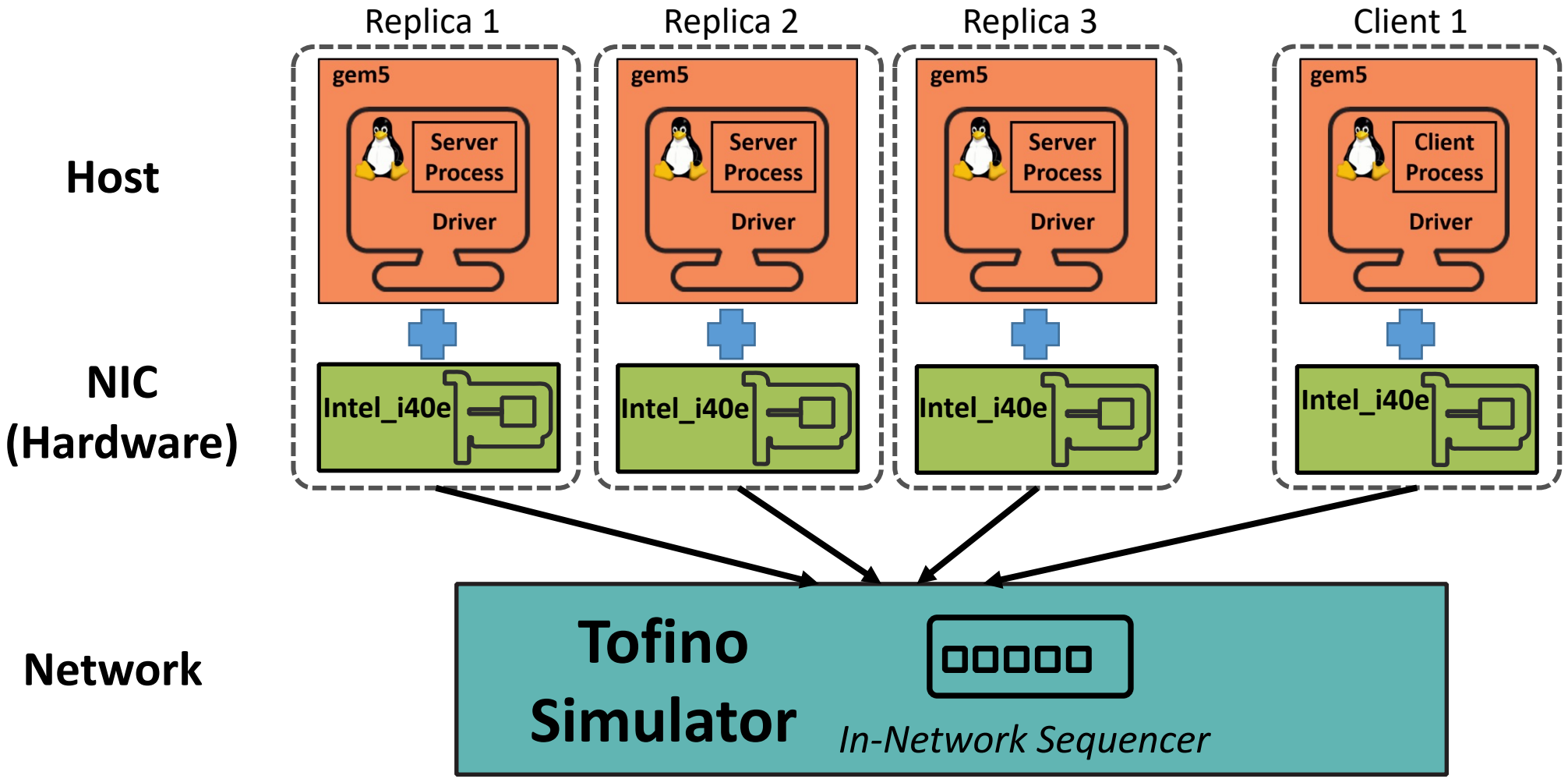
- Inline synchronization with messages
- Pairwise synchronization is sufficient
- Link latency provides slack



Evaluation Overview

- **End-to-End:** simulate full systems including HW & SW
- **Scalable**: simulate large systems with tens or hundreds of devices
- **Fast**: keep simulation times as low as possible
- **Modular**: enable flexible “plug & play” composition of simulators
- **Accurate**: preserve accuracy of simulators, correctly interface and synchronize them.
- **Deterministic**: keep simulations deterministic when components are deterministic and synchronized

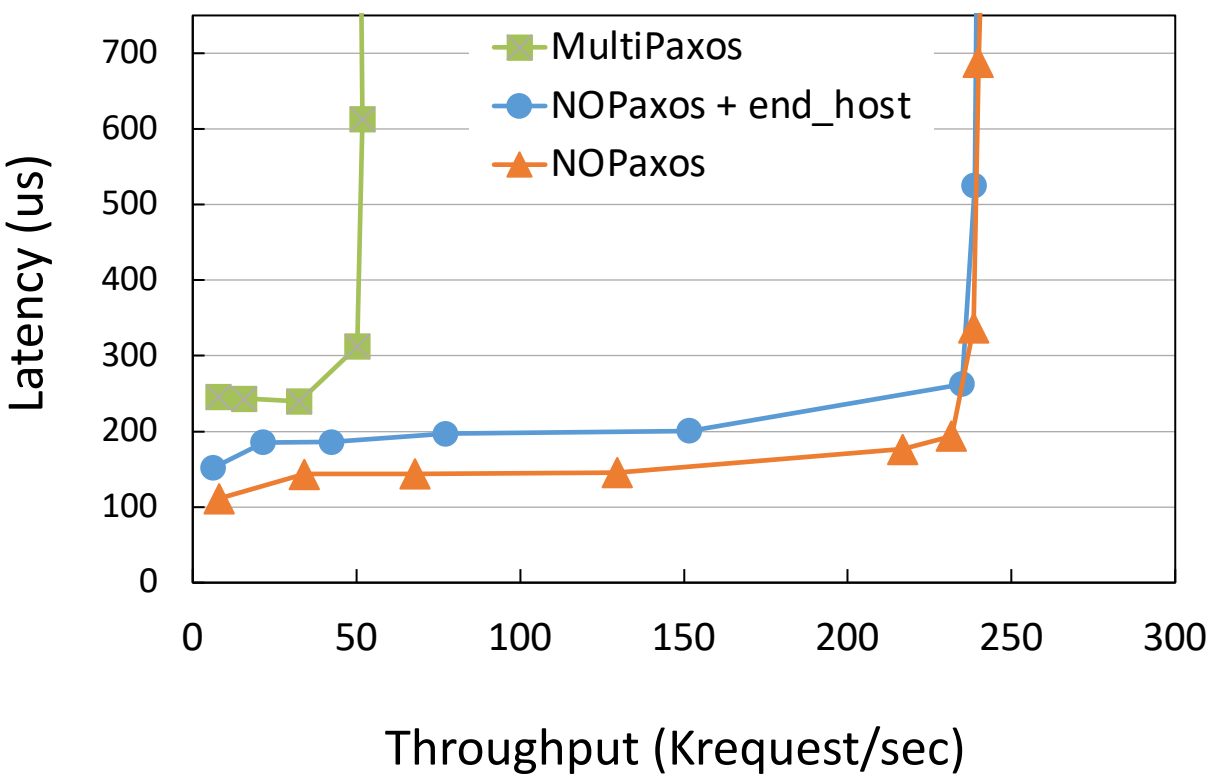
NOPaxos E-to-E System Evaluation



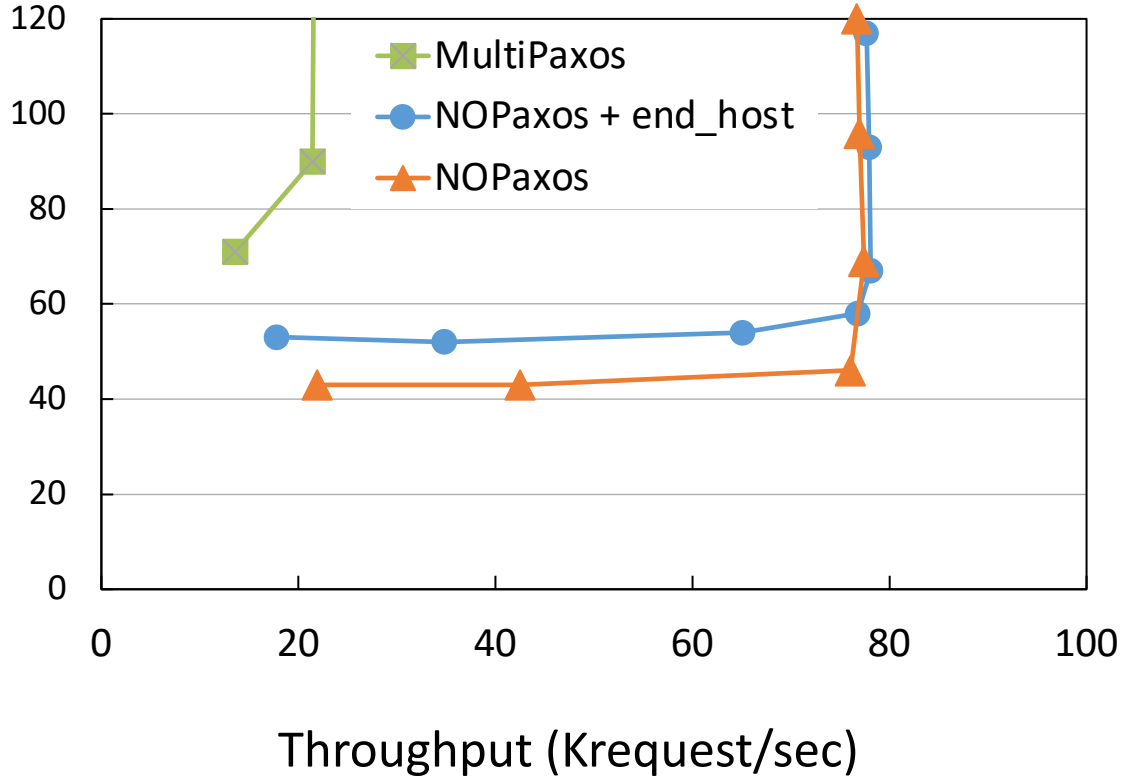
NOPaxos E-to-E System Evaluation

- SimBricks can reproduce the main properties

Original Result From the Paper

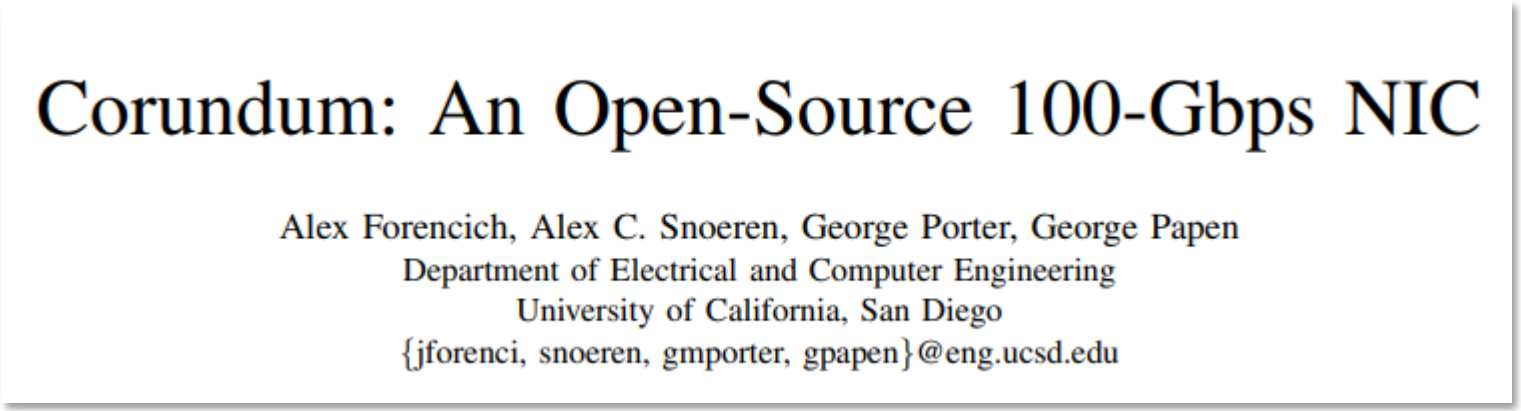


Simulated Result with SimBricks



Corundum NIC E-to-E System Evaluation

End-to-end evaluation with unmodified FPGA RTL and drivers



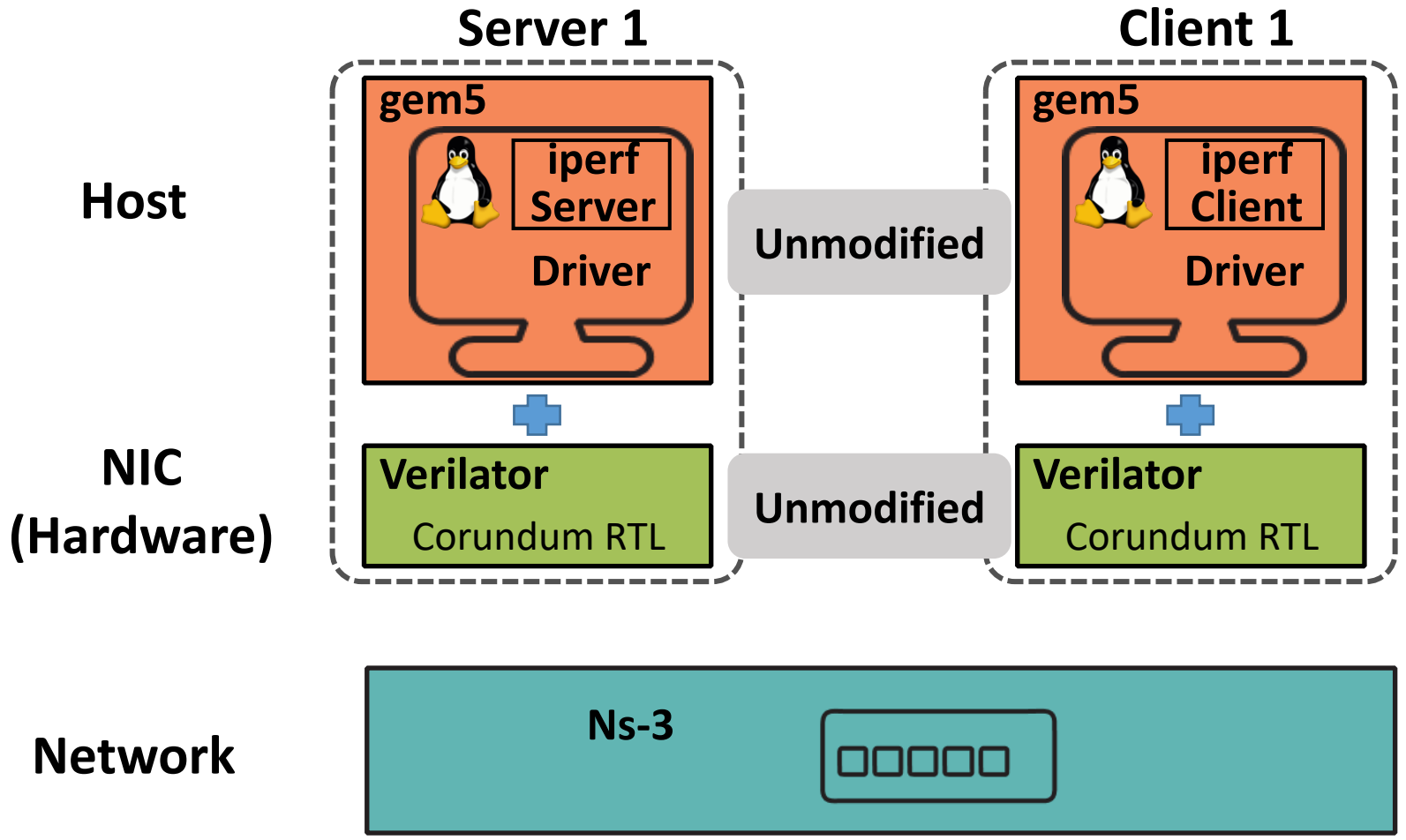
Corundum: An Open-Source 100-Gbps NIC

Alex Forencich, Alex C. Snoeren, George Porter, George Papen
Department of Electrical and Computer Engineering
University of California, San Diego
{jforenci, snoeren, gmporter, gpapen}@eng.ucsd.edu

[FCCM'20]

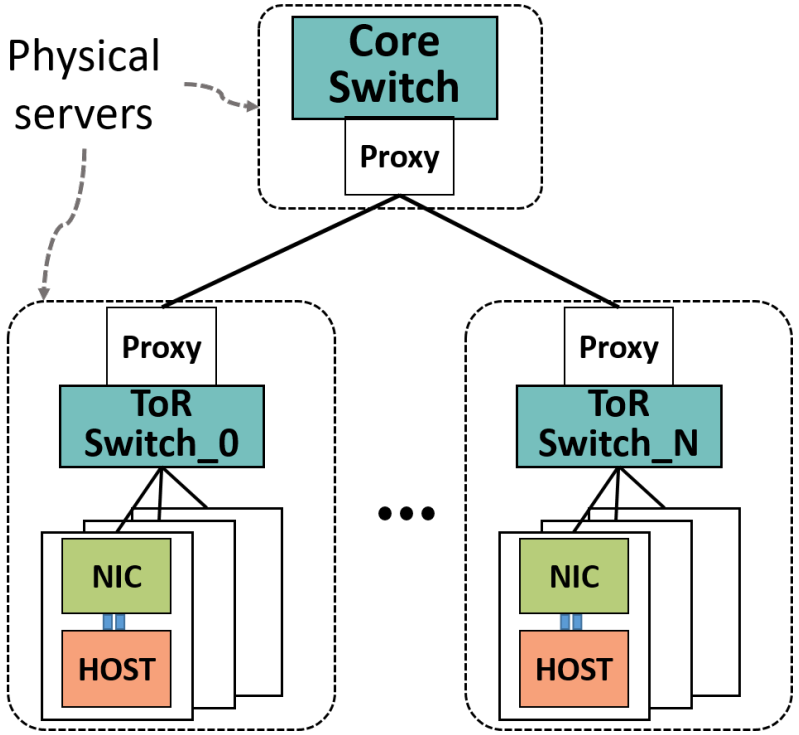
Corundum NIC E-to-E System Evaluation

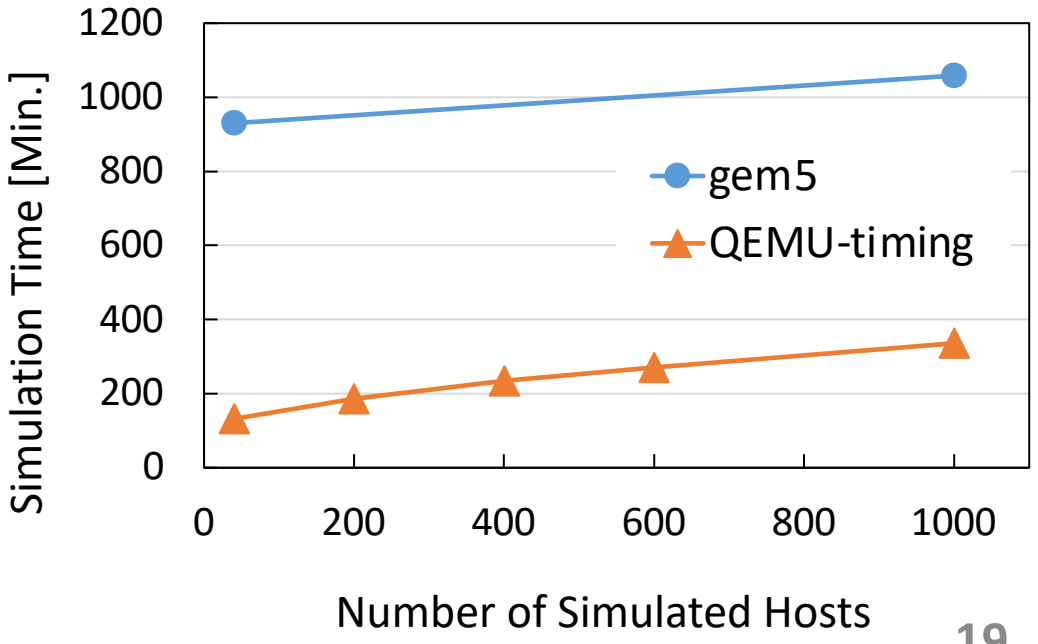
End-to-end evaluation with unmodified FPGA RTL and drivers



SimBricks is Scalable

- Scales to simulate systems up to 1000 hosts
- Proxy: Spans simulation to different physical hosts





Summary

- **SimBricks** combines component simulators into full systems
  - Synchronization for meaningful performance results
- End-to-end evaluation of full post-Moore systems
  - Capable of running unmodified SW & HW systems
  - Side-benefit: reproducible results with deterministic simulators!

Host	gem5	QEMU		
H/W	Intel i40e Behavioral	Corundum Verilator	Corundum Behavioral	FEMU (SSD)
Net	Tofino	ns-3	OmNet	L2 Switch

https://github.com/simbricks

https://simbricks.github.io