Architecting to Achieve a Billion Requests Per Second Throughput on a Single Key-Value Store Server Platform

Sheng Li[1], Hyeontaek Lim[2], Victor W. Lee[1], Jung Ho Ahn[3], Anuj Kalia[2], Michael Kaminsky[1], David G. Andersen[2], Seongil O[3], Sukhan Lee[3], Pradeep Dubey[1]

[1]Intel Labs, [2]Carnegie Mellon University, [3]Seoul National University

Important and challenging In-memory Key-value Stores

In-memory distributed key-value store

- Critical data caching/serving layer in cloud/datacenter infrastructure
- Low-latency, high-throughput, simple hash-table-like interface, at scale

Red hot research in two directions

- Hardware/FPGA acceleration, w/ stock software
- Software centric research, w/ stock hardware





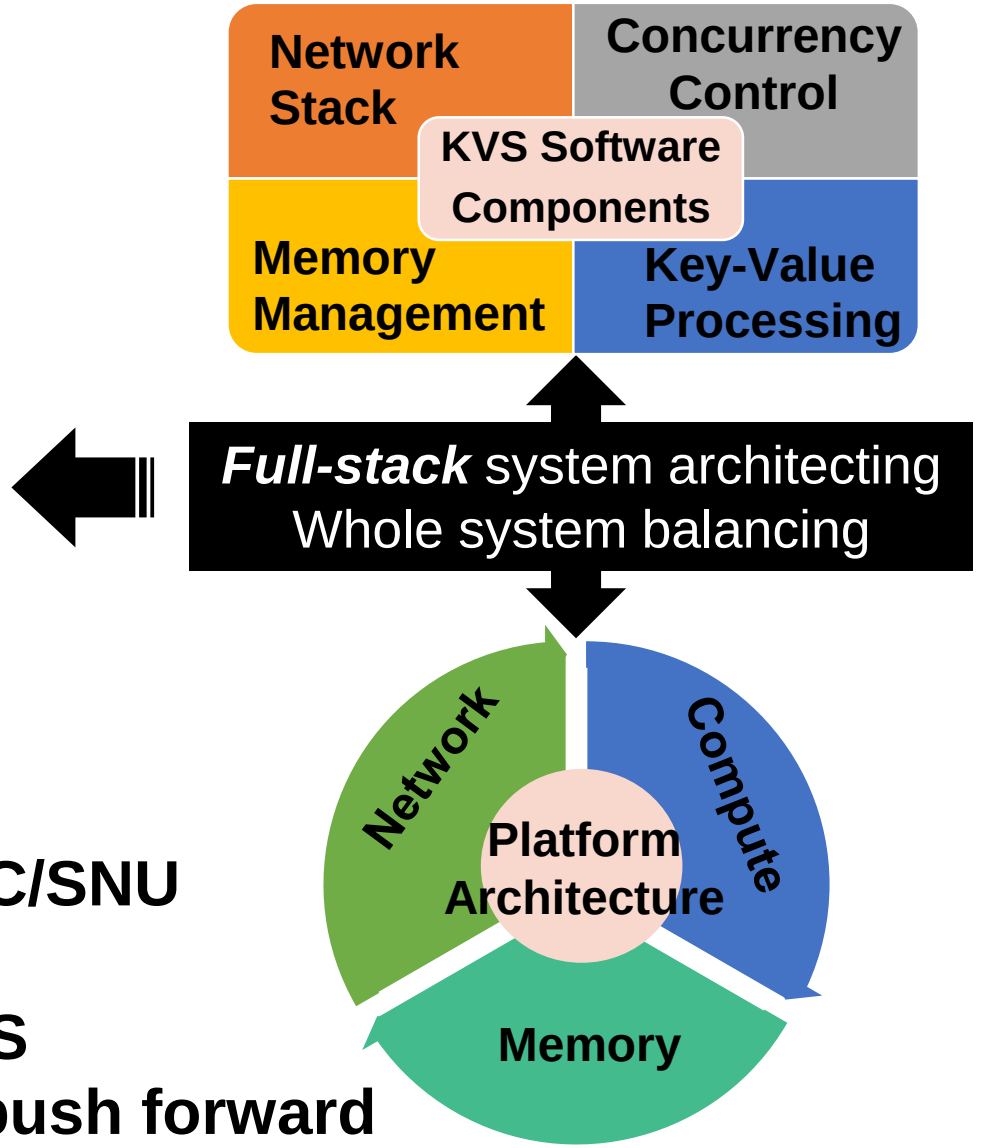





Understanding the essence of key-value stores

Achieving 120 Million requests/second (RPS) on commodity hardware

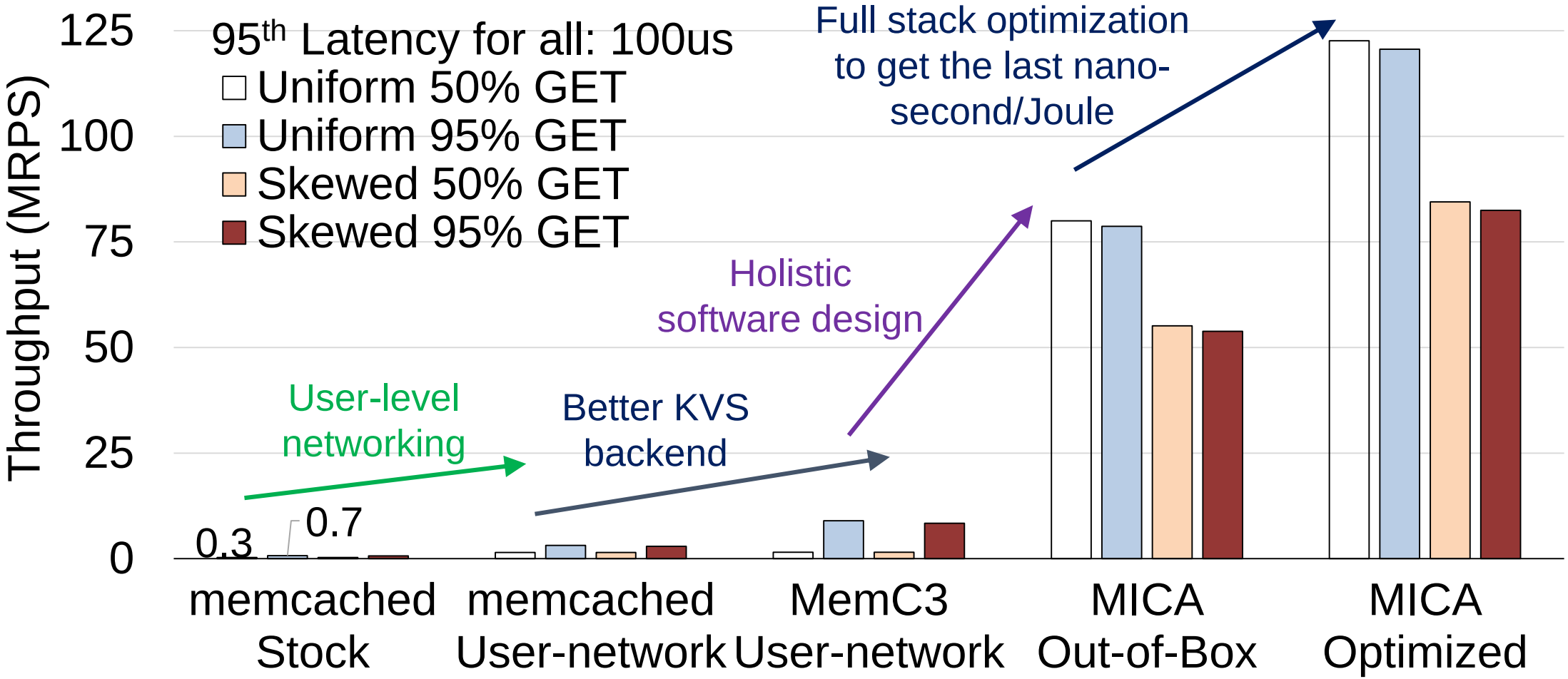
Architecting a near future KVS platform achieving 1 Billion RPS and beyond



Great collaboration between Intel Labs/ISTC-CC/SNU

- **Published in ISCA' 2015**
- **Honored by being fast-tracked to ACM TOCS**
- **Strong interests from IL, DCG, and SSG to push forward**

Record-setting Performance & Energy Efficiency

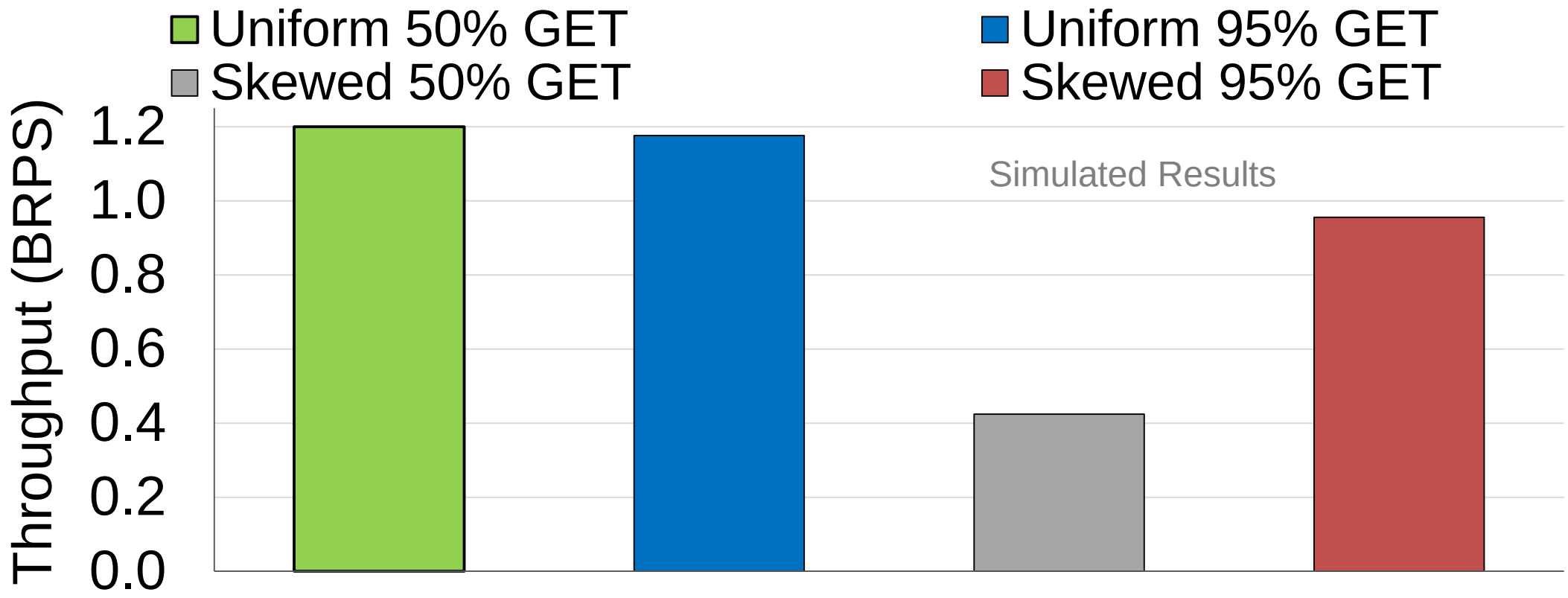


Best energy efficiency among all systems (FPGA, ARM, and GPU)

We are No.1
We are not done

The pursuit of another 10X speedup

Future Platforms: to Billion RPS and Beyond



Quad-socket with 60 3-issue cores, 750KB L2 cache/core, 300GbE per socket, with flow-director and DDIO

Full-stack architecting as the key enabler

- Optimized MICA
- Manycore based, whole system optimized, balanced platform

Conclusions

Understanding the essence of key-value stores

Achieving 120 Million requests/second (RPS) on commodity hardware

Architecting a near future KVS platform achieving 1 Billion RPS and beyond

Great collaboration between Intel Labs/ISTC-CC/SNU

- **Published in ISCA' 2015**
- **Honored by being fast-tracked to ACM TOCS**
- **Strong interests from IL, DCG, and SSG to push forward**