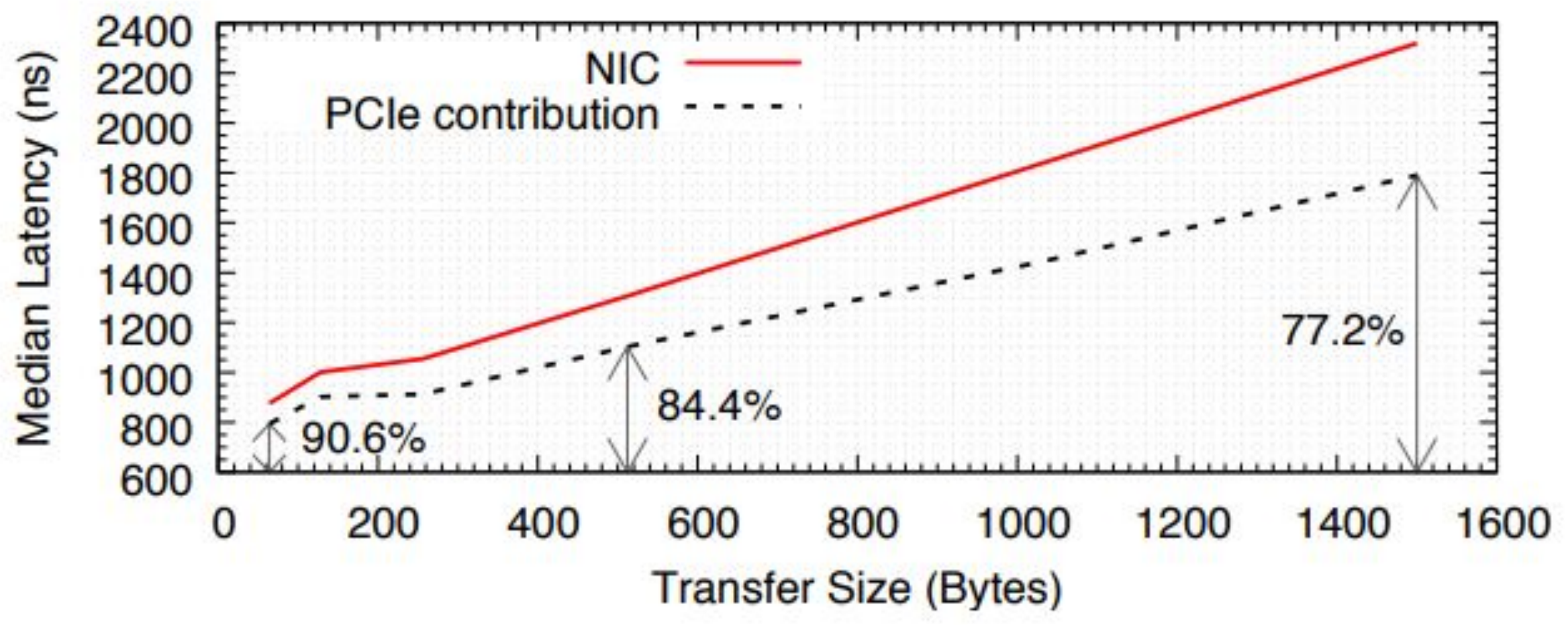
CCI-Bench: Measuring Cache-Coherent Interconnection (Part of Project Optimus for FPGA Virtualization)

Gefei Zuo, Jiacheng Ma

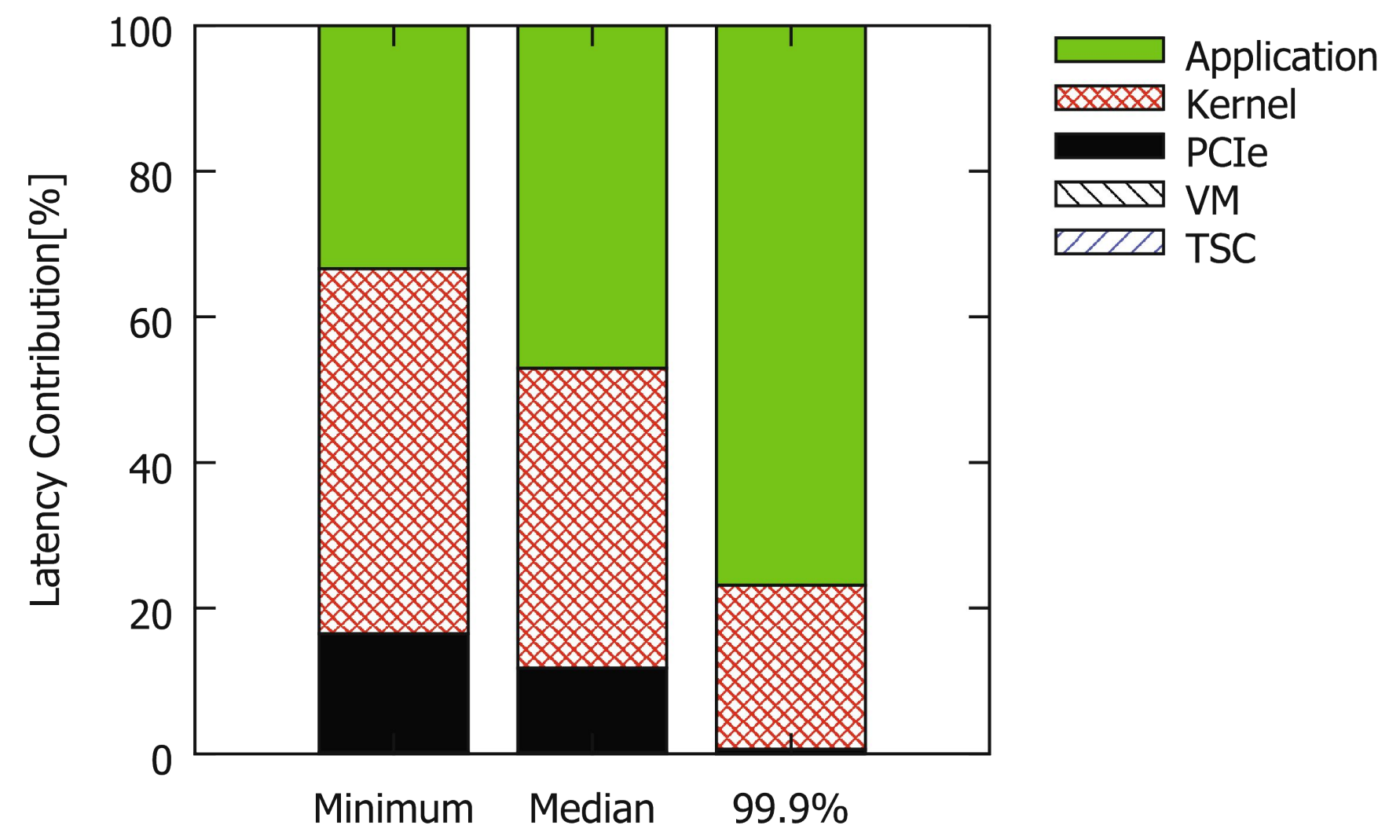
Why Interconnection is important to NIC?

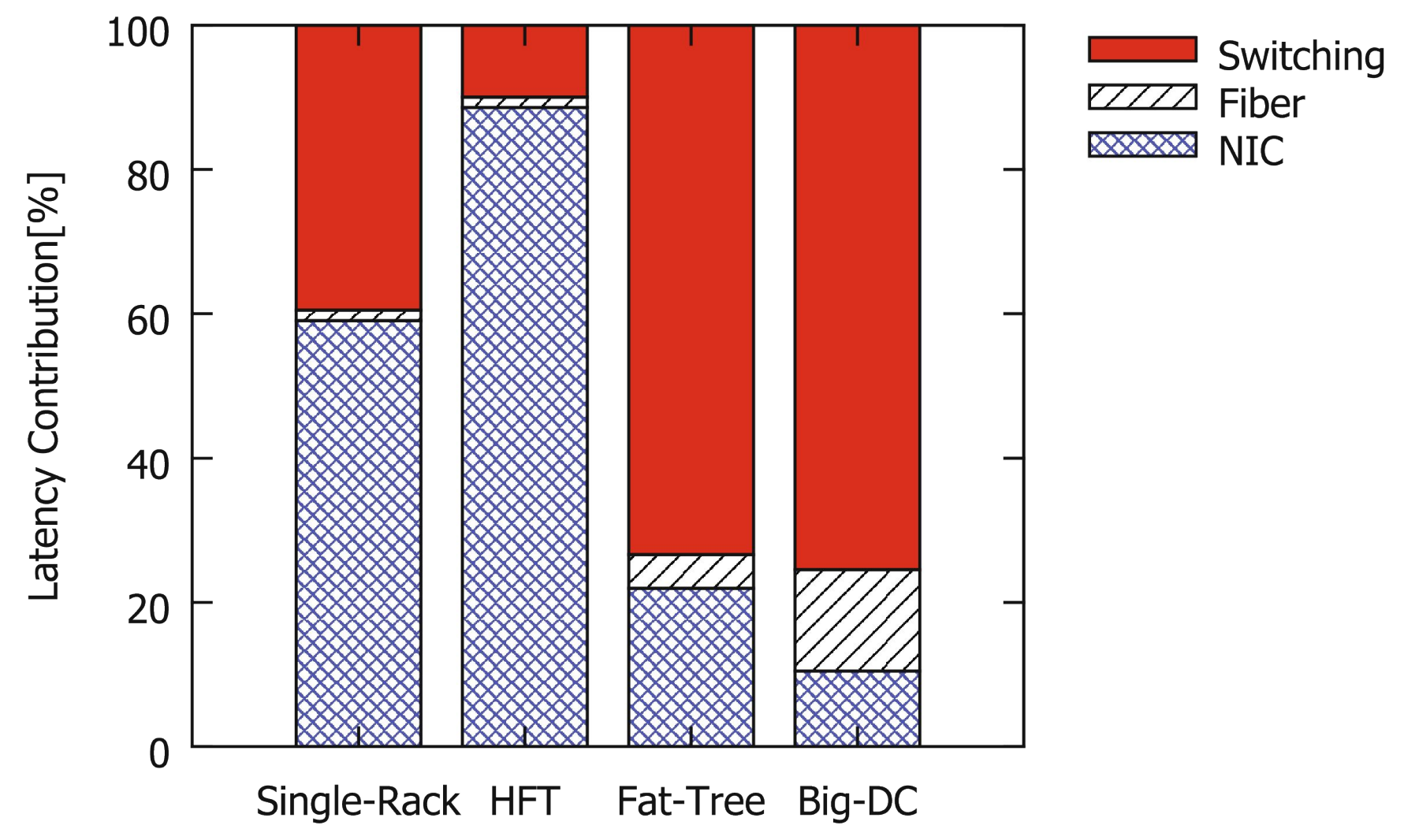
- It contributes a non-trivial portion to network latency, especially for small packets



[1] R. Neugebauer, G. Antichi, J. F. Zazo, Y. Audzevich, S. López-Buedo, and A. W. Moore, "Understanding PCIe Performance for End Host Networking," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, New York, NY, USA, 2018, pp. 327–341

- It's an unexplored optimization opportunity.
  - It's important under simple network topology
  - Application optimization, new congestion control, kernel-bypass network stack, won't reduce interconnection latency.

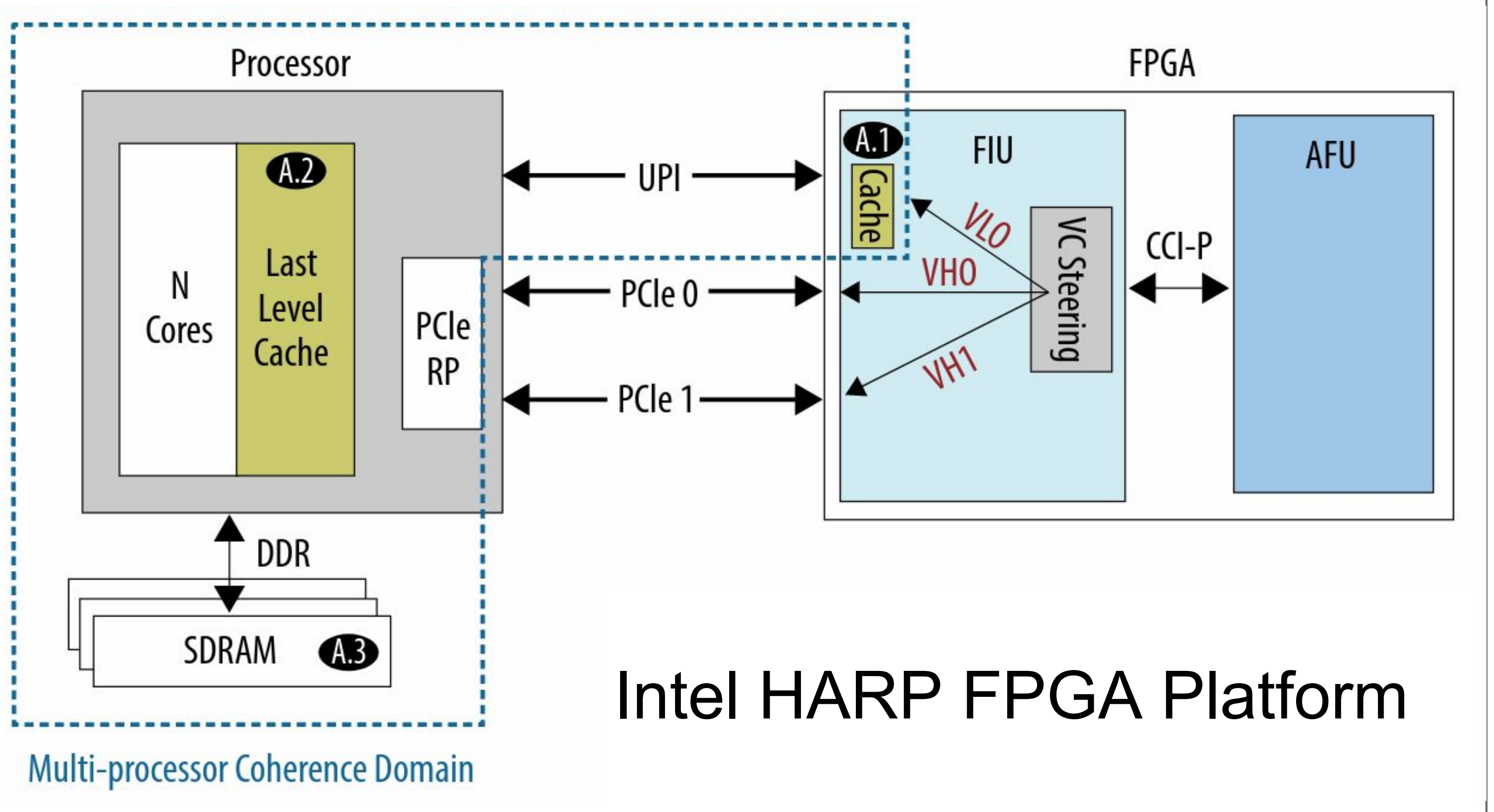




[2] N. Zilberman *et al.*, "Where Has My Time Gone?," in *Passive and Active Measurement*, 2017, pp. 201–214.

Cache-Coherent Interconnection

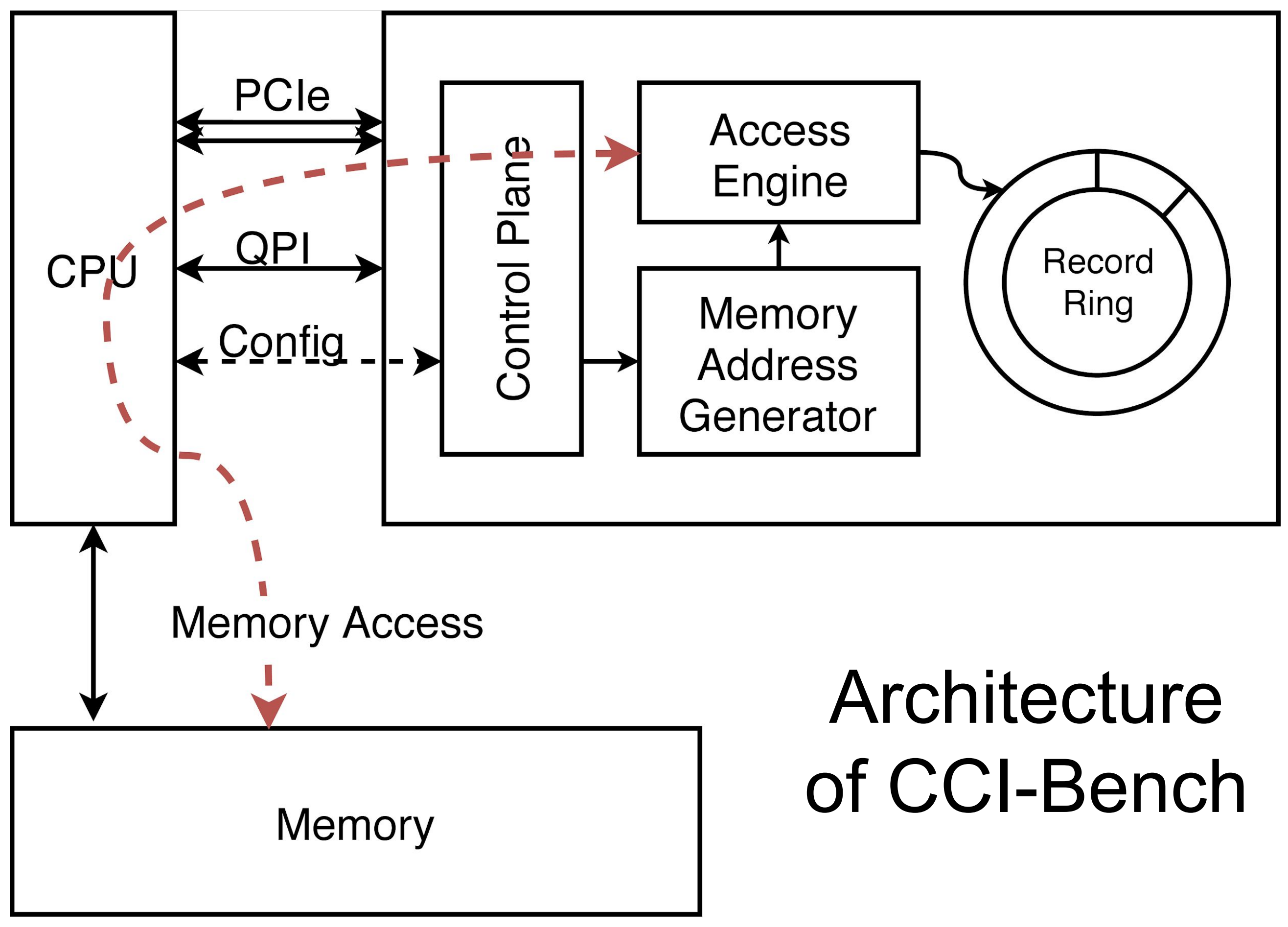
- Inter-CPU communication
  - e.g. Intel UPI and QPI
  - connects different NUMA nodes
- High bandwidth & low latency, a complementary of PCIe
- Ideal protocol to connect NIC to the CPU in the future



Intel HARP FPGA Platform

How to measure?

- Intel HARP FPGA
- Integrated CPU+FPGA
- Two versions:
  - HARPv2: QPI
  - HARPv3: UPI
- **CCI-Bench**
  - RTL implemented benchmark for memory access measurement

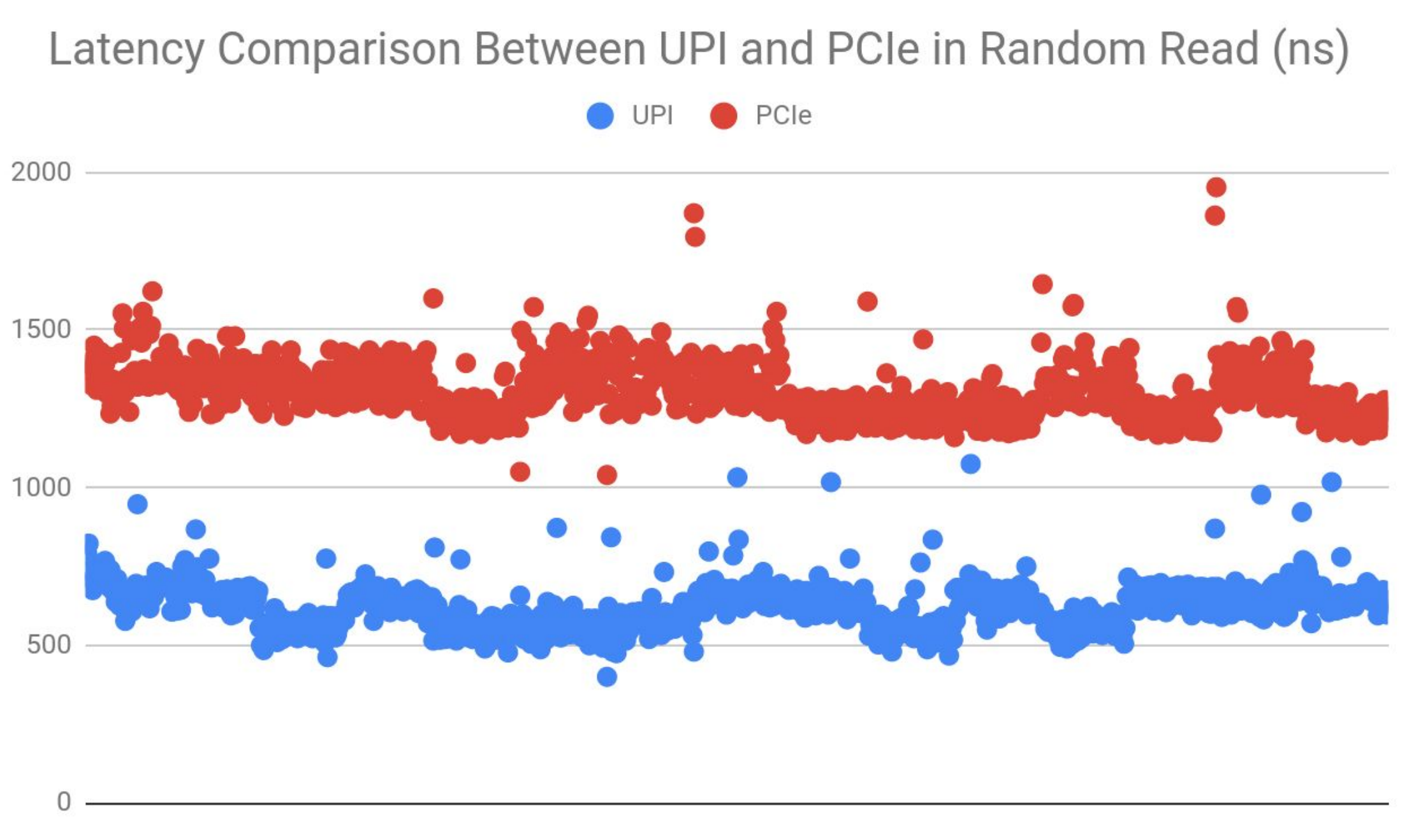


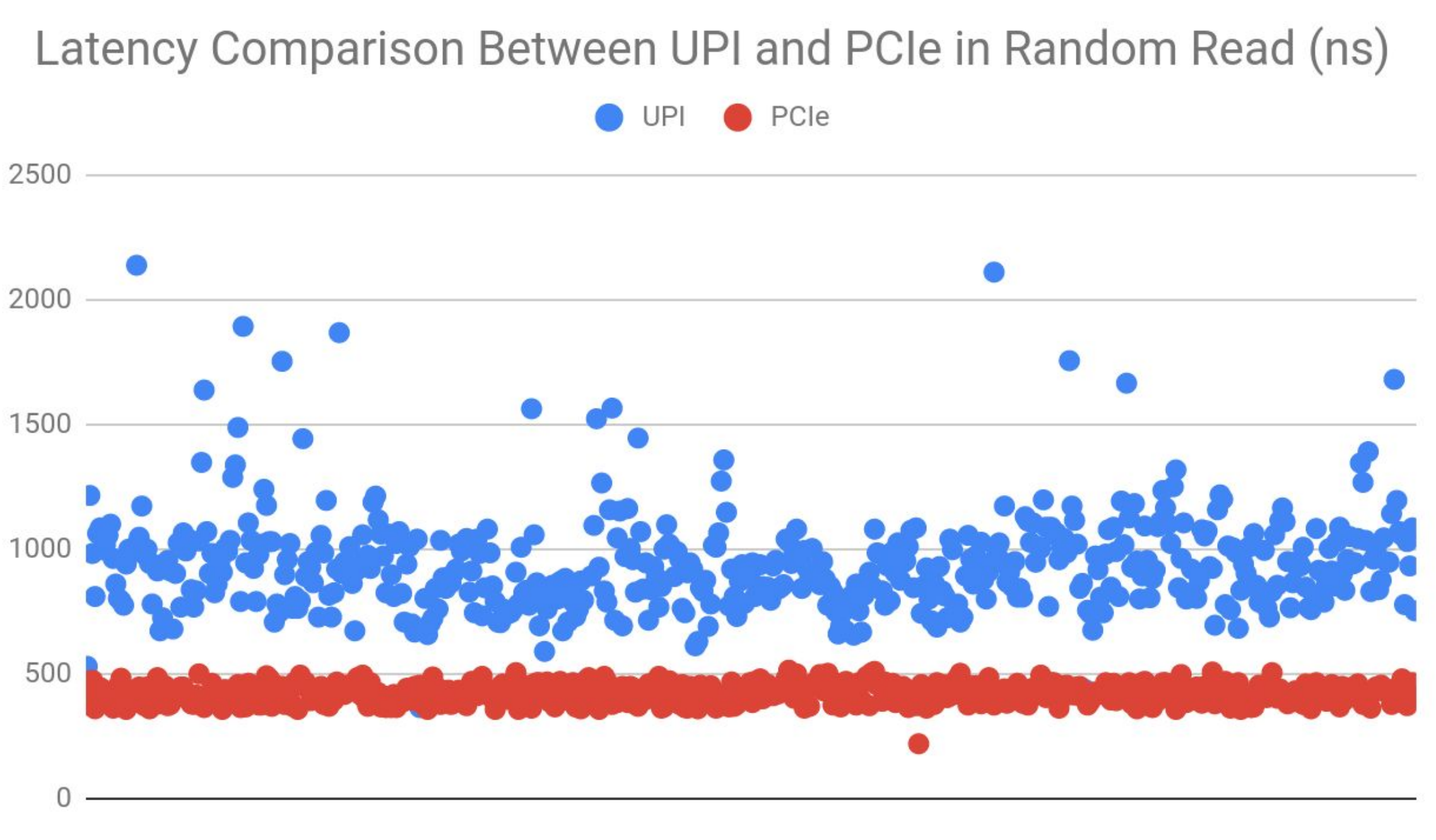
Architecture of CCI-Bench

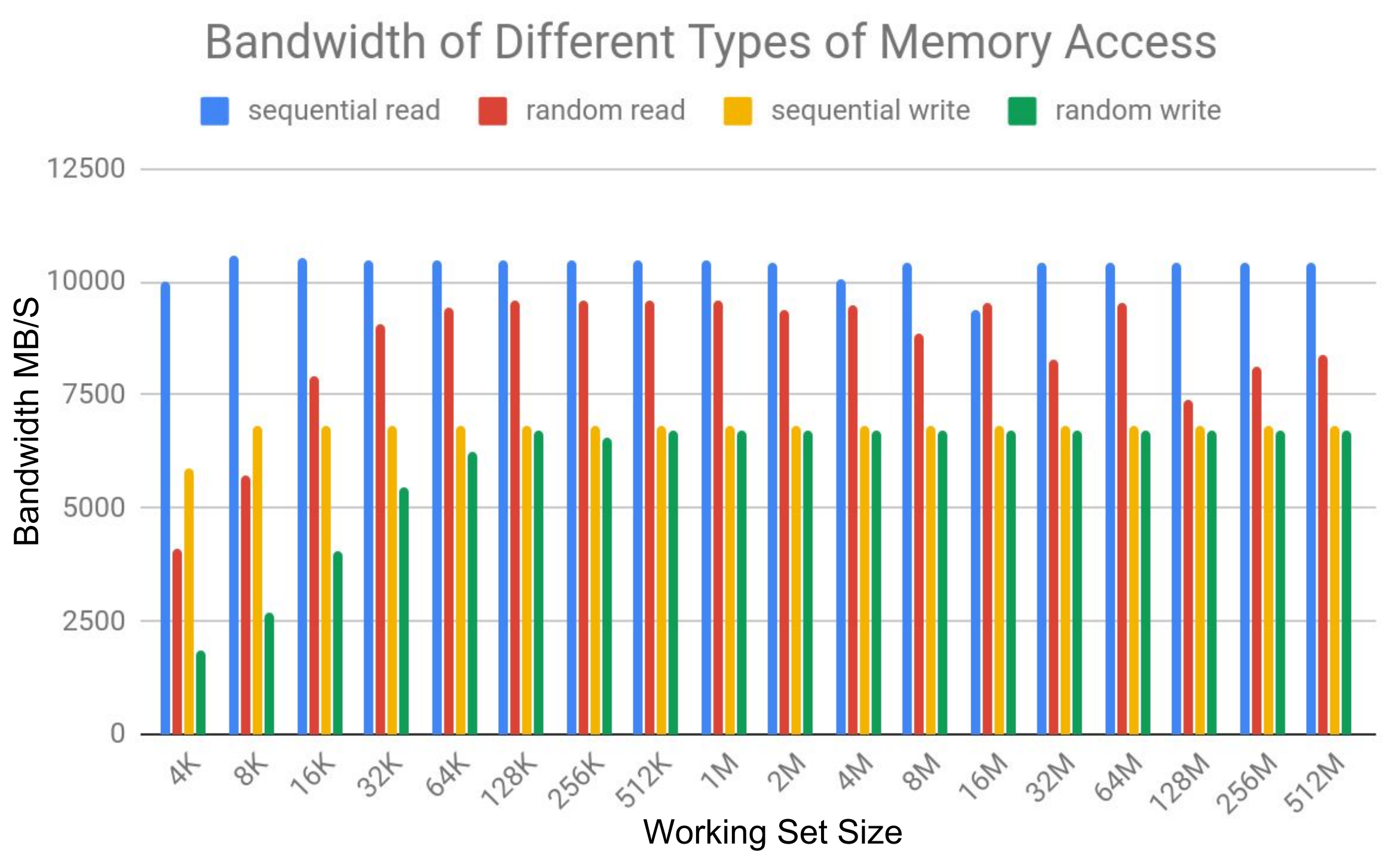
CCI-Bench

- Perform memory access to a configurable range in a configurable manner
- Software Part:
  - Allocate memory
  - Configure the FPGA
- Hardware Part:
  - Generate addresses and access them
  - Record latency in a ring buffer

Measurement Result

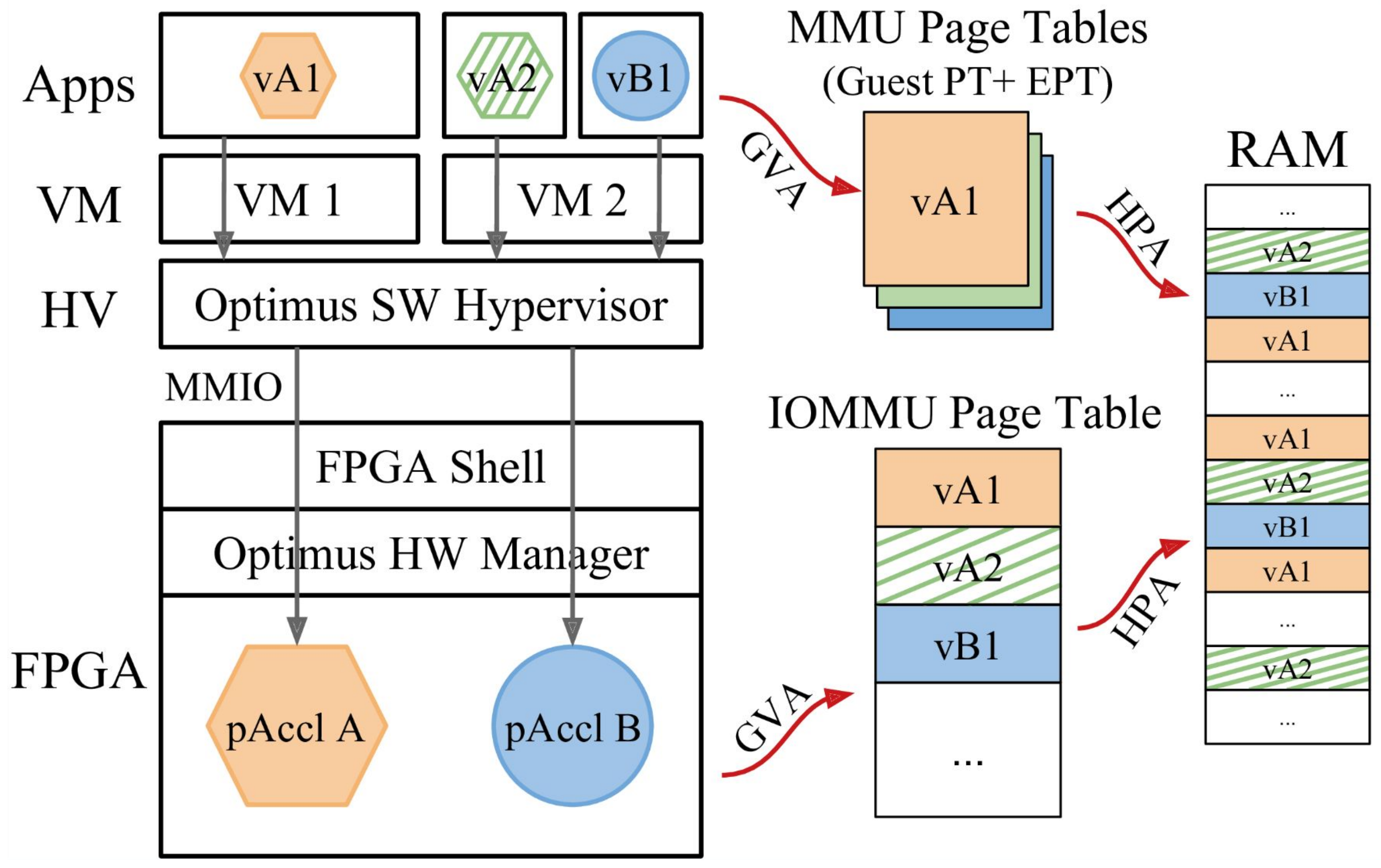






About Project Optimus

- The first hypervisor for shared memory FPGA platforms (e.g. HARP)
- MMIO access are trapped and emulated
- DMA bypasses the virtualization layer with the help of IOMMU
  - Page Table Slicing (each VMs own part of the page table)
  - Enabling virtualization without SR-IOV or other hardware support
- Spatial Multiplexing + Temporal Multiplexing
- Scalability:
  - Spatial: 9 Accelerators
  - Temporal: as much VMs as you like


Apps
vA1
vA2
vB1
VM
VM 1
VM 2
HV
Optimus SW Hypervisor
MMIO
FPGA Shell
Optimus HW Manager
FPGA
pAccl A
pAccl B
MMU Page Tables
(Guest PT+ EPT)
GVA
vA1
HPA
RAM
...
vA2
vB1
vA1
...
vA1
vA2
vB1
vA1
...
vA2
...
IOMMU Page Table
vA1
vA2
vB1
...
GVA
HPA