

ExaNeSt Project

- **Interconnects**

- Low-latency, unified compute & storage traffic

- **Storage**

- Fast, distributed, in-node non-volatile memory

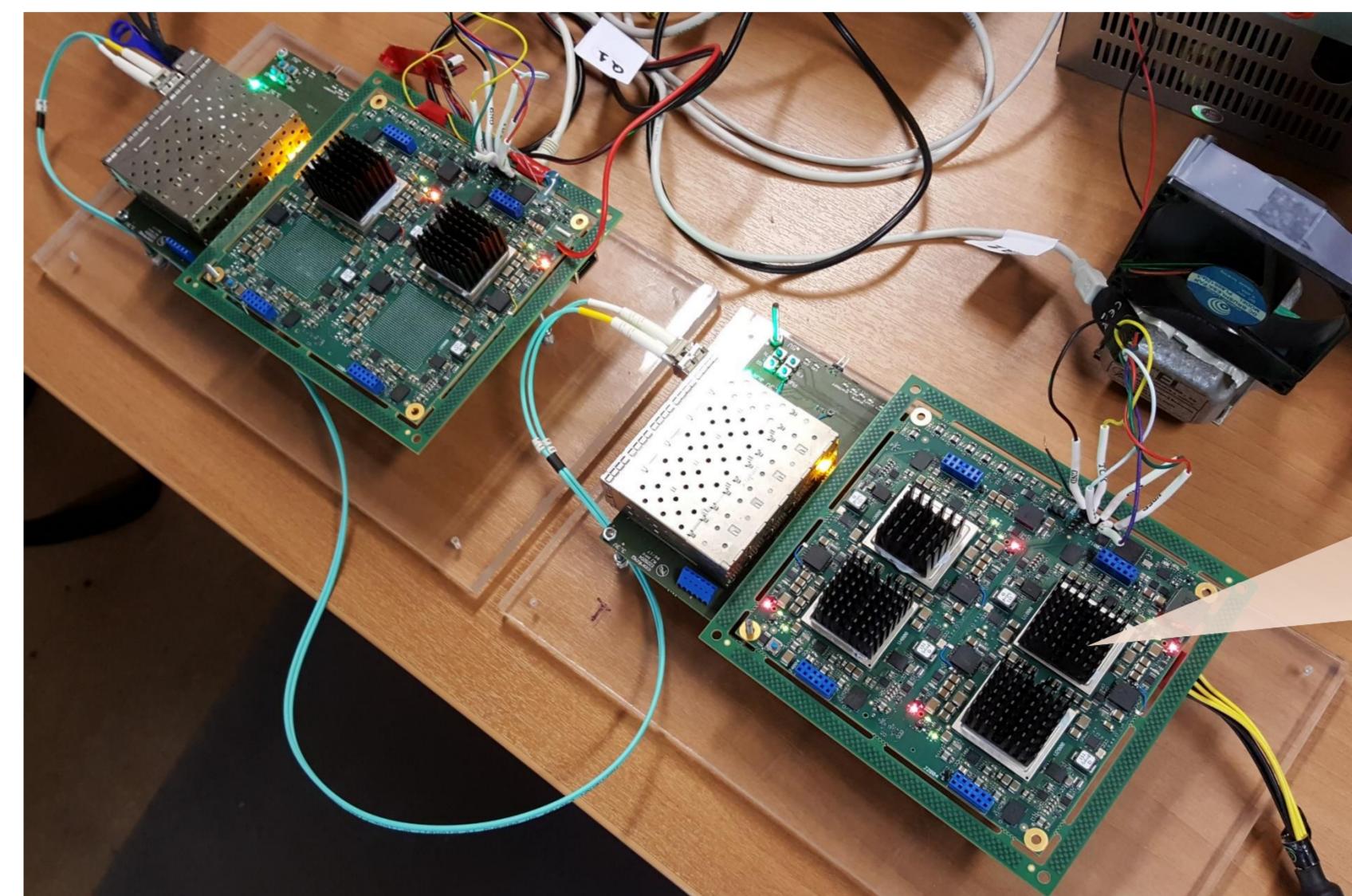
- **Applications**

- Real applications: scientific computing, data analytics

- **System Packaging Technology**

- Compact, fully-immersed liquid cooling technology

ExaNeSt Prototype based on QFDB



Quad FPGA Daughter board
(16 64-bit ArmV8 cores)
60 Watt

- 4 Xilinx Zynq Ultrascale+ MPSoCs (4 64-bit ARM cores) + 64 GB DRAM + 250GB SSDs
- ~80 liquid-cooled QFDBs in final prototype

Minimizing Communication Latency

- **Network Interface**

- Virtualized 8-channel DMA engine
- Virtualized 16-channel ARM I/O MMU (SMMU)

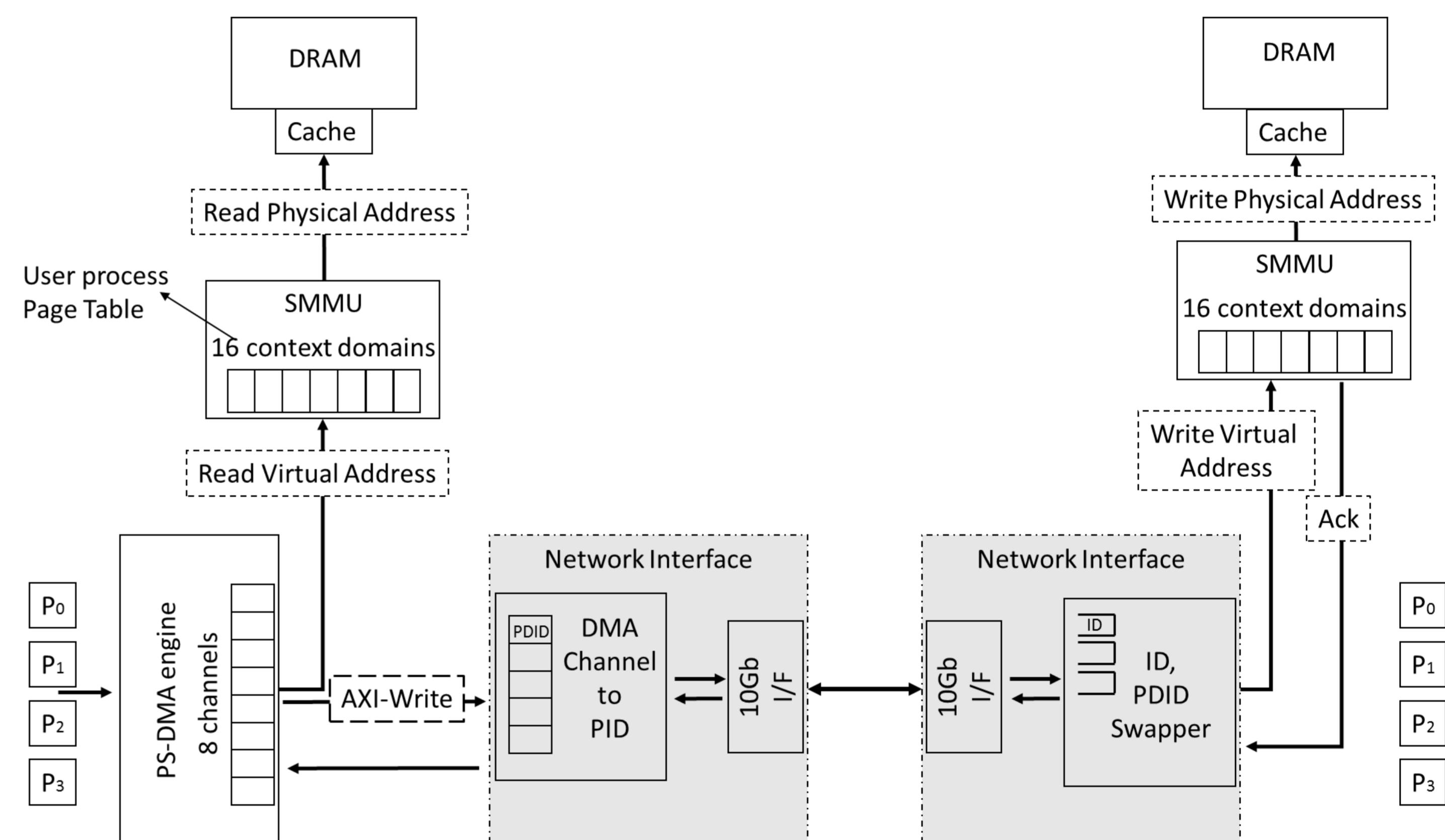
- **Global Virtual Address Space**

- Addresses are translated at destination node's SMMU
- No page pinning, no extra copies at destination

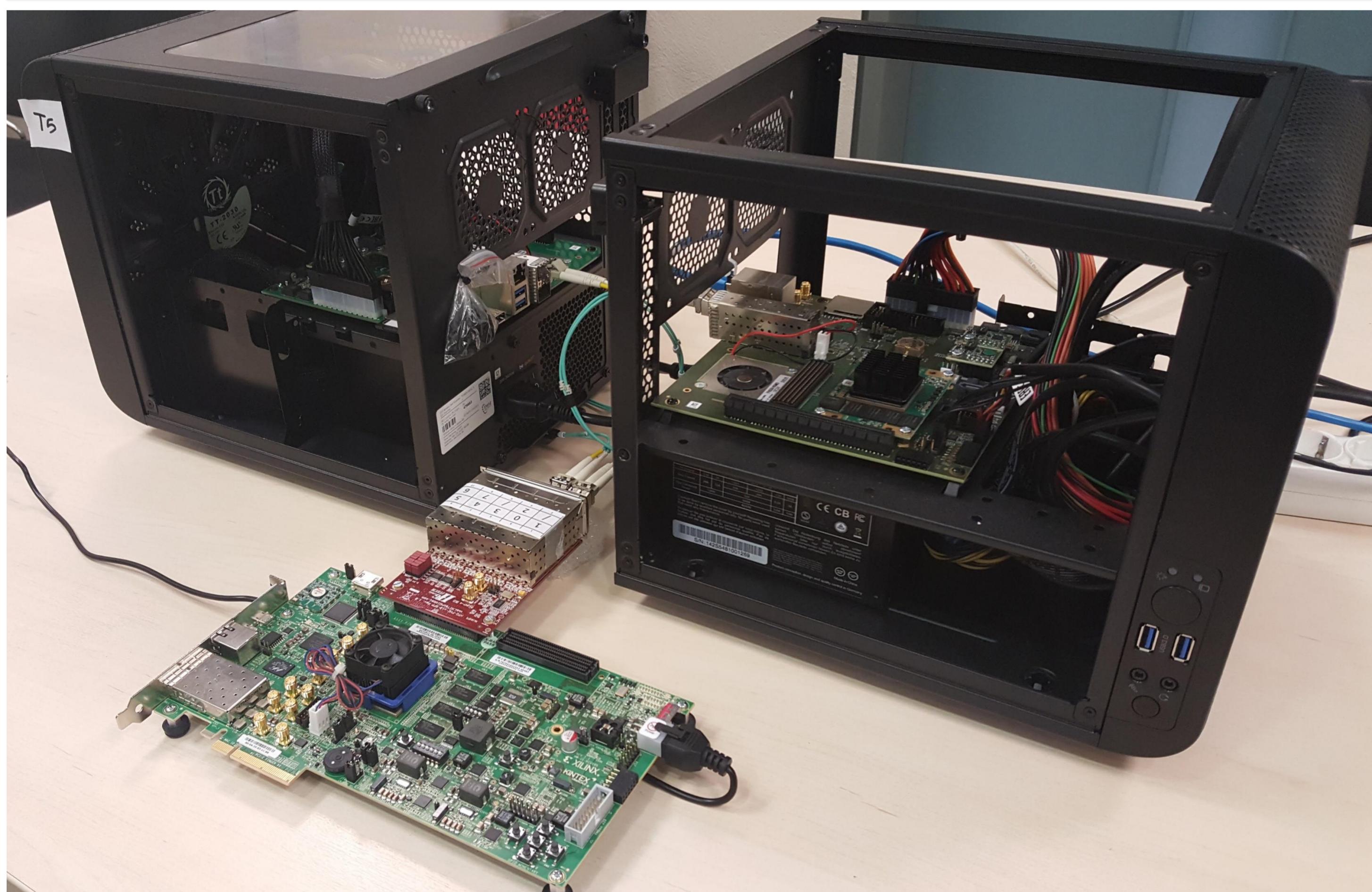
- **User-level Initiated RDMA**

- Kernel bypass to reduce latency

End-to-end RDMA flow

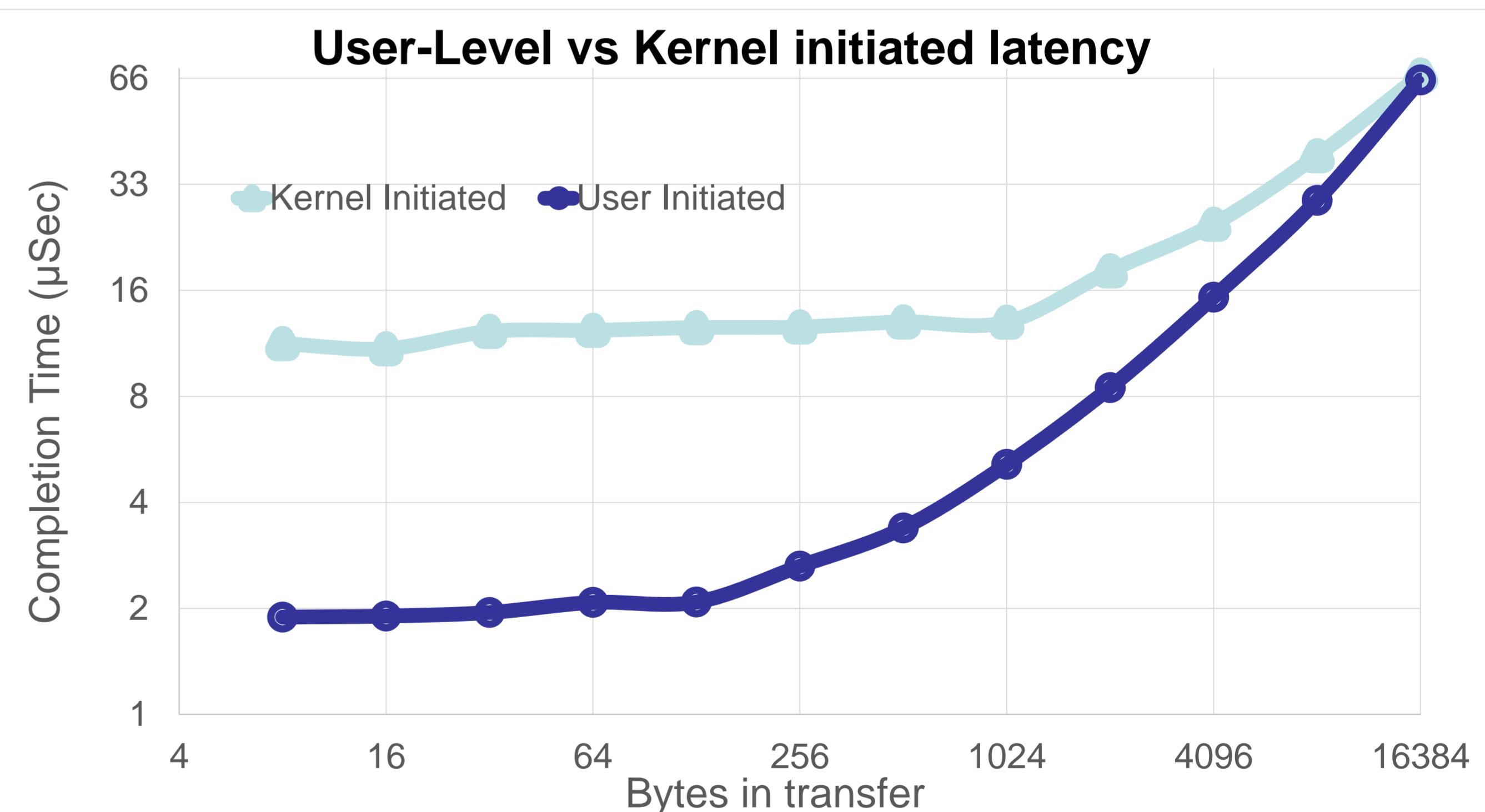


Experimental prototype



- Commercial Zynq Ultrascale+ MPSoC boards (Trenz)
- Interconnect: Xilinx Kintex UltraScale used as central AXI switch, providing 10 Gb/s SFP+ links

Evaluation



Conclusion and future work

- 2 μsec RDMA latency (8x speedup wrt. kernel-initiated transfers)
- virtualized user-level access to packetizers and MBOXes
- Next items
 - porting MPI
 - new low-latency ExaNeSt interconnect
 - fast completion notification & congestion control