An FPGA-based Prototyping Platform for Research in High-Speed Interprocessor Communication

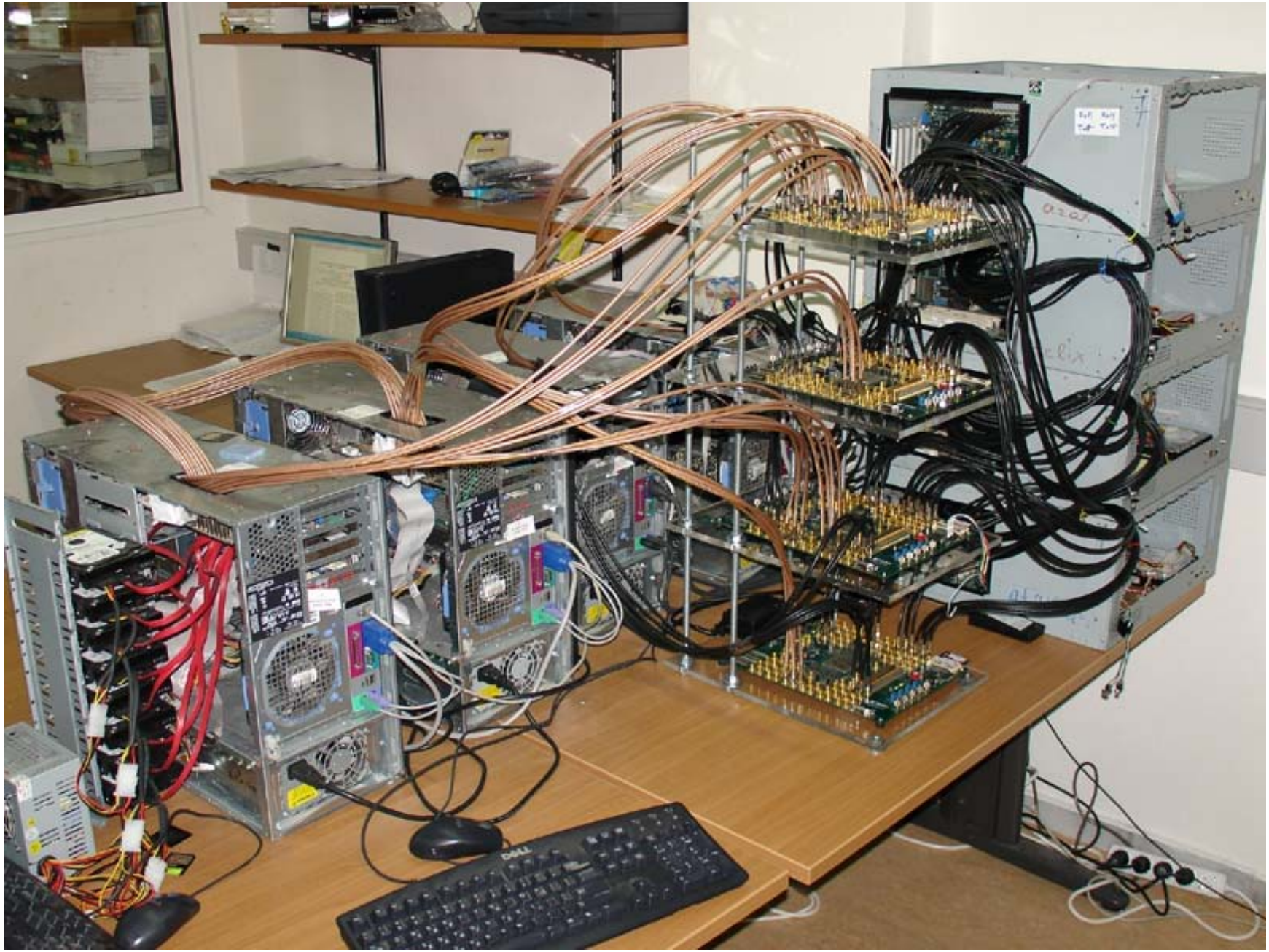
V. Papaefstathiou, G.Kalokairinos, A.Ioannou, M.Papamichael, G.Mihelogiannakis, S.Kavadias, E.Vlahos,

D.Pnevmatikatos and M.Katevenis

Inst. of Computer Sci. (ICS) – FORTH – Crete, Greece

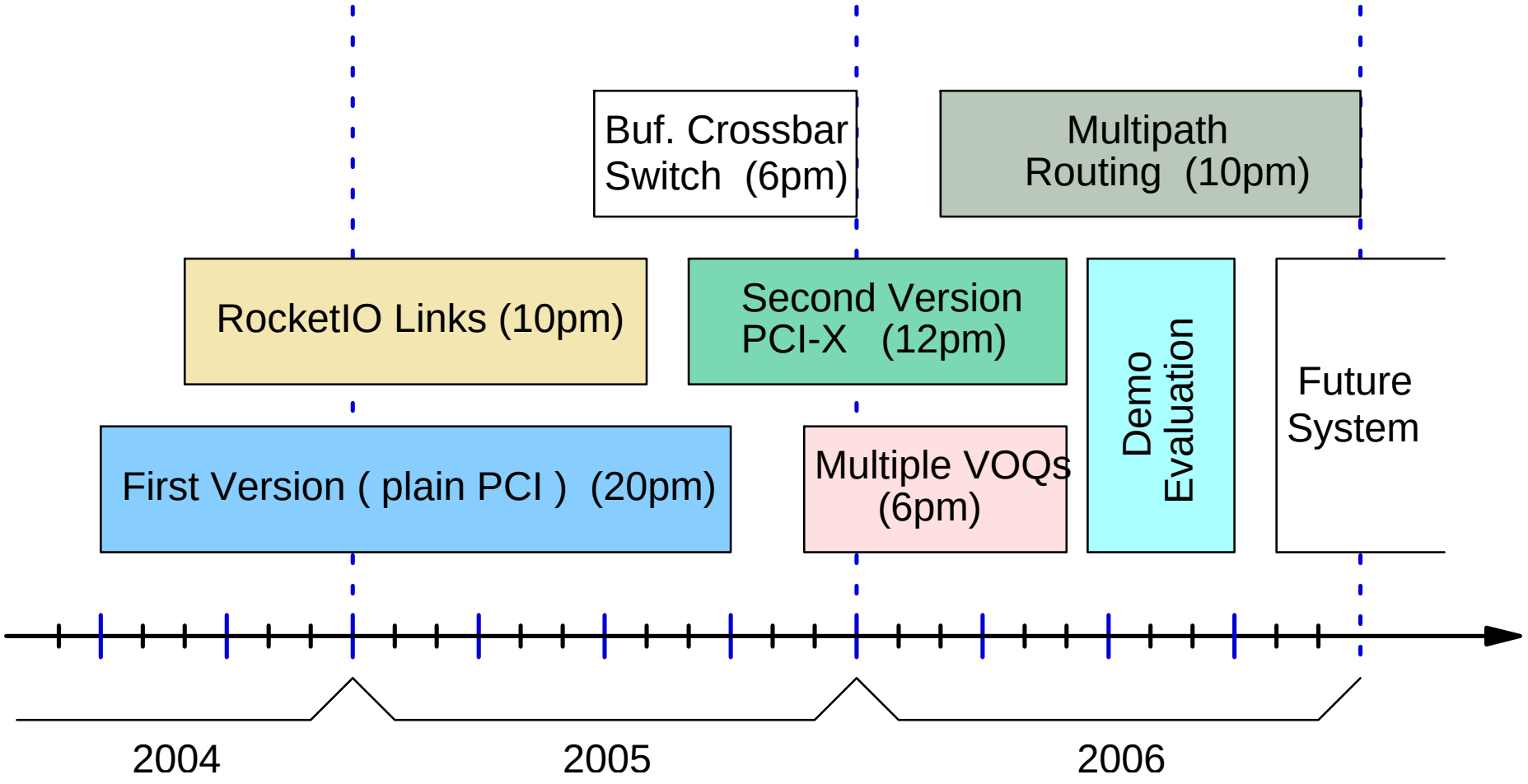
Key Features (First System)

- Remote DMA (RDMA) based operation
- Notifications: departure and/or arrival, interrupt or enqueue
- Remote Enqueue for short messages, multiple senders
- Credit-based flow control: lossless communication
- Per-destination Virtual Output Queues (VOQ's): flow isolation
- Extensive event logging, debugging & performance counters
- Switch:
  - 8x8 implemention – 32-bit datapath @78.125 MHz
  - achieved up to 16x16 – 16-bit dpth @156.25MHz to fit in FPGA
- Linux already adapted for this platform (kernel-mode comm.)
- MPI port for this platform under way



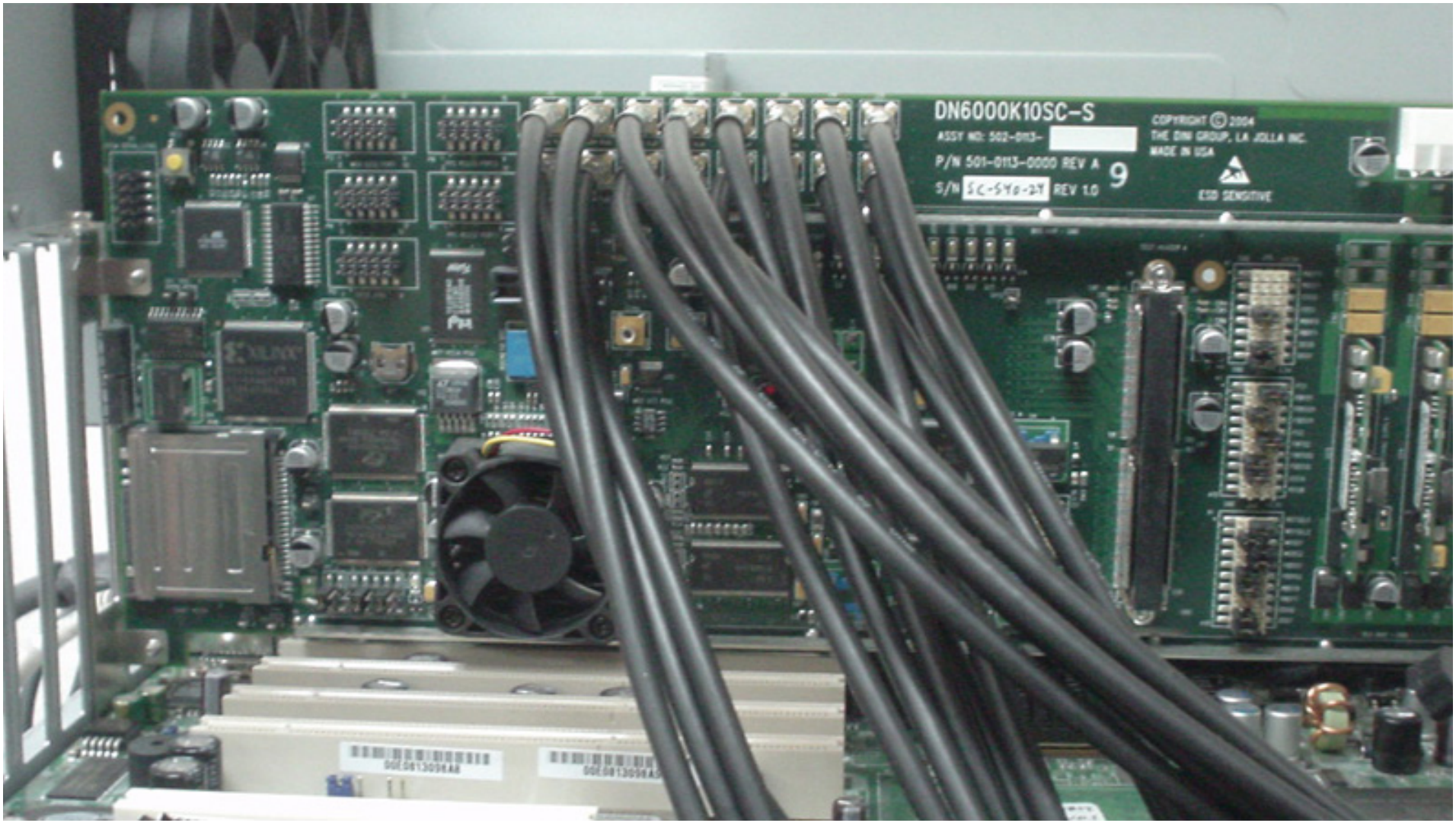
Photograph of First System (8 nodes)

Hardware Development Cost (1st System, 2 versions)



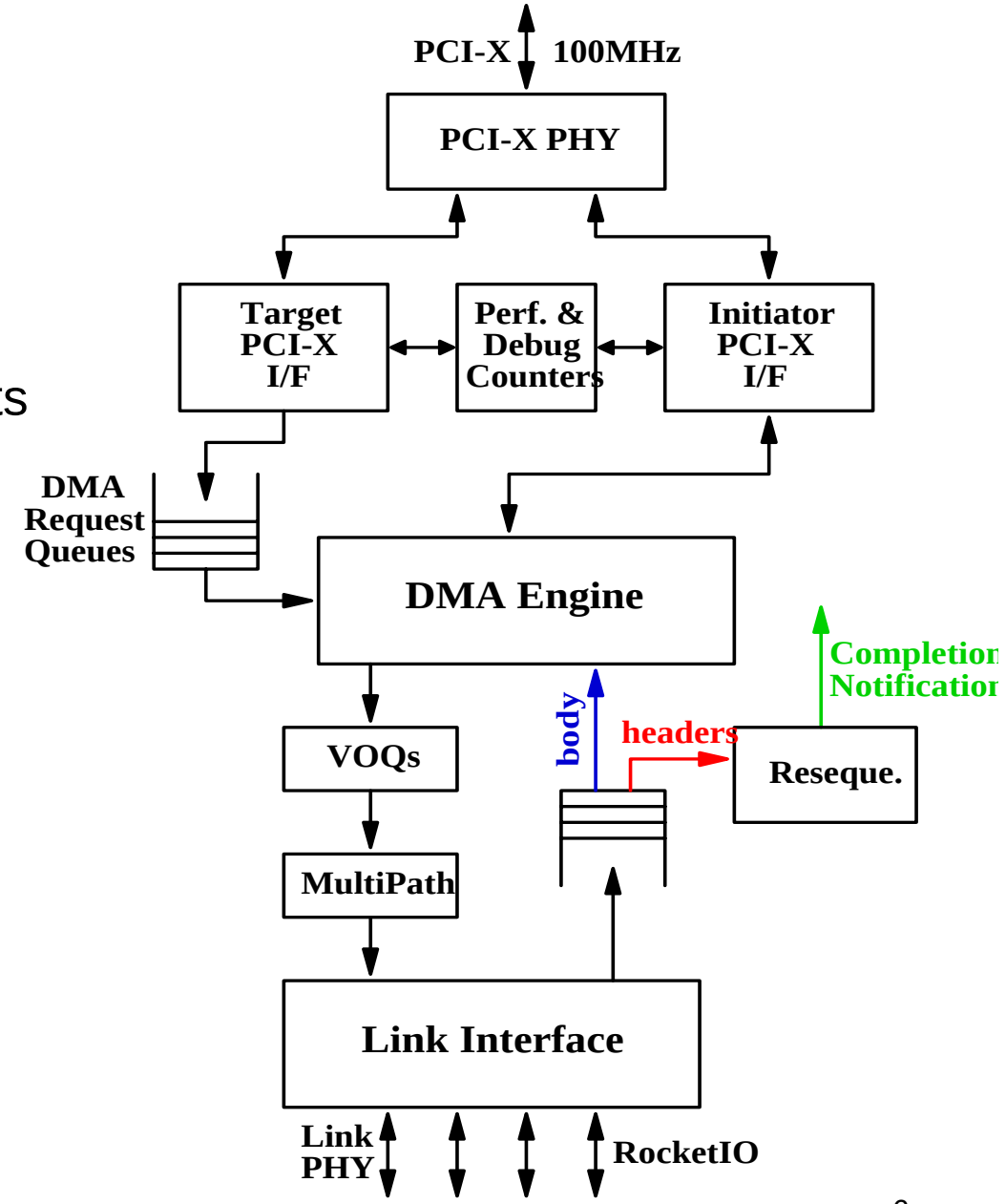
pm = person-months

NI Photo, with 4 RocketIO links

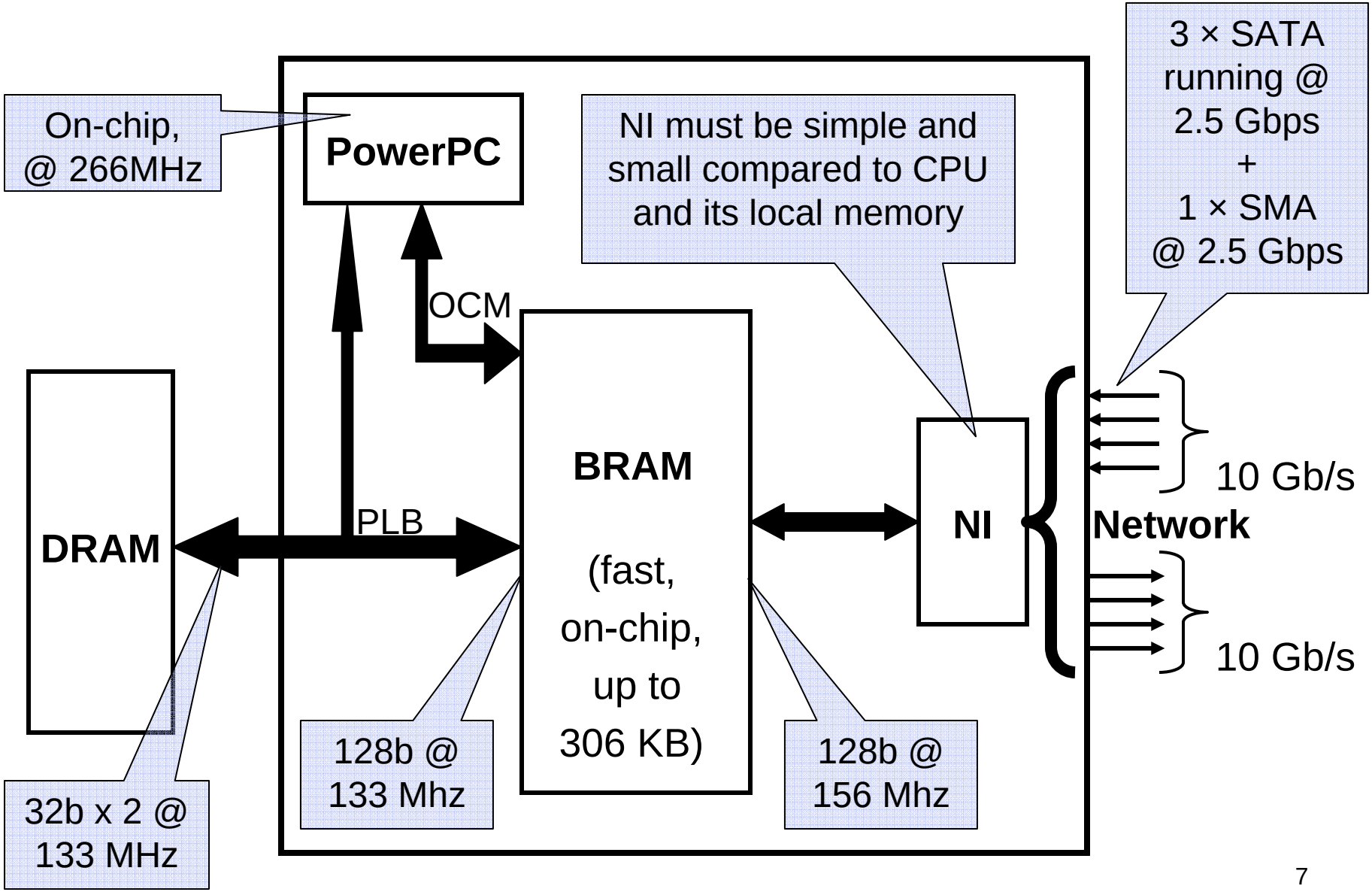


NI Architecture

- 64-bit PCI-X @100MHz to Host
- Per-dest. DMA request Q's
- Per-destination VOQ's
- DMA segementation into packets
- Link bundling: 4×2.5 = 10 Gb/s
  - inverse multiplexing
    multipath routing
- Out-of-order packet arrivals:
  - DMA body immediately written into memory
  - headers wait in resequ. Q's
  - nxt sys: just count # bytes
  - notify completion after resequ. / count complete
  - resequ. tolerates 1-pck loss



Next Generation (2007) Node: Block Diagram



Next-Generation System Node (Xilinx Univ.Pr.)