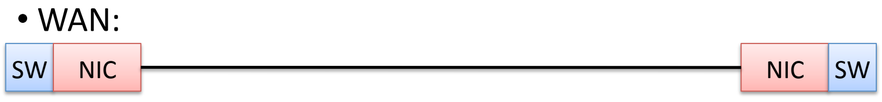
Authors: Mario Flajslik, Mendel Rosenblum

Title: Low Latency Network Interfaces

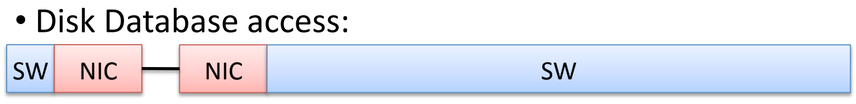
motivation

Request – Response Latencies

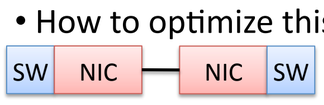
- WAN:



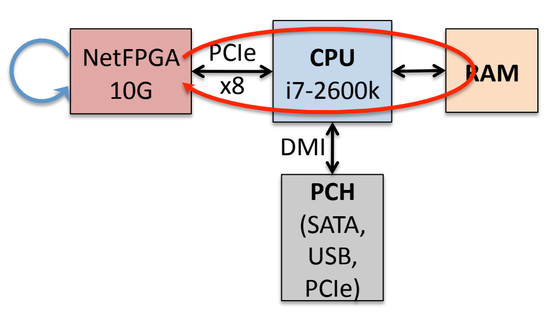
- Disk Database access:



- How to optimize this?:



Test Setup

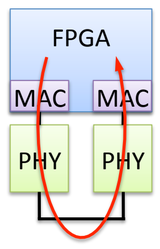


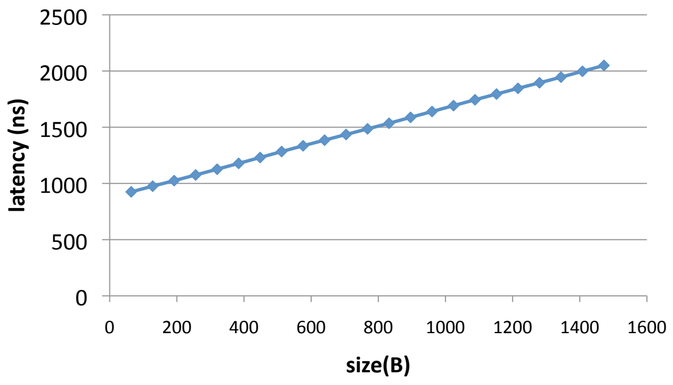
Wire Latencies

Cable	Speed	Latency
Coax	0.65c	5.13 ns/m
Twisted Pair	0.59c	5.65 ns/m
Optical	0.66c	5.05 ns/m
Twinax	0.65c	5.10 ns/m

10G Ethernet

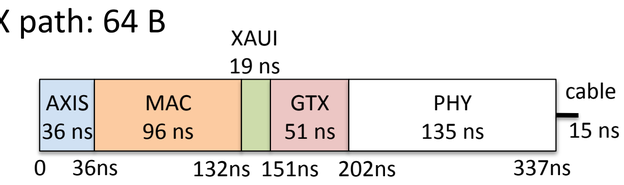
10G Ethernet (3m twinax cable)



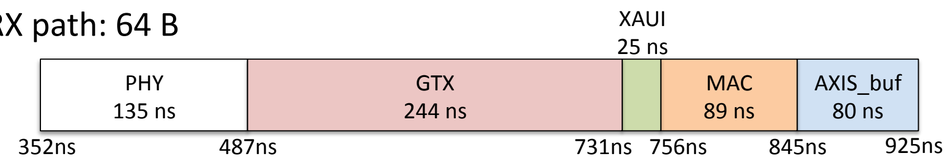


Latency breakdown

- TX path: 64 B

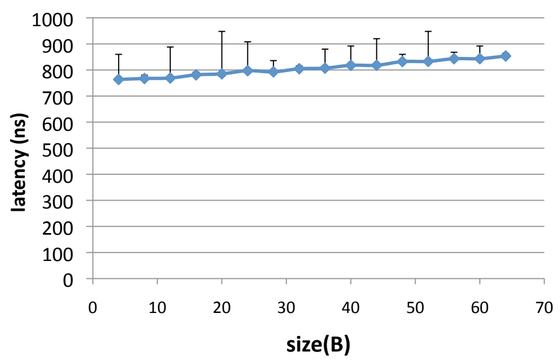


- RX path: 64 B



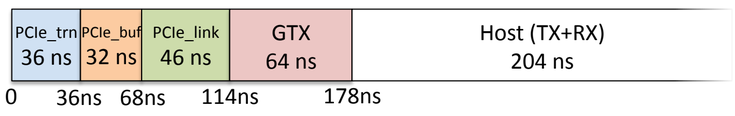
PCIe

PCIe Read Latency

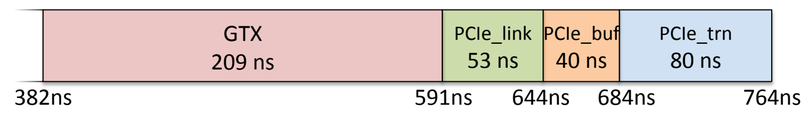


PCIe Latency Breakdown

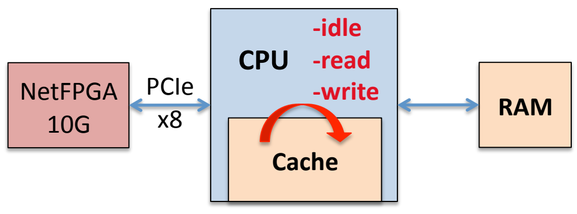
- TX path: 12 B



- RX path: 16 B



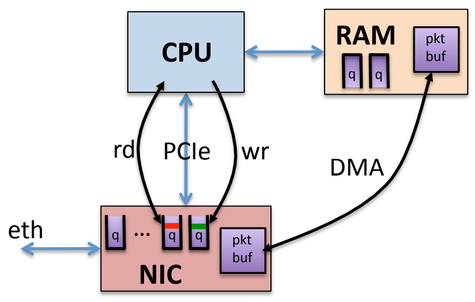
Cache Effects



PCIe 4B read	CPU idle	CPU read	CPU write
average	768 ns	768 ns	756 ns
stddev	1.90 ns	2.02 ns	8.59 ns
min	764 ns	764 ns	736 ns
max	776 ns	784 ns	772 ns

Interface

PCIe Read Latency



Best case 60B packet
TX+RX latency: 2us

60B packet transmit example

```
if(credits_available()){
    credits_dec();

    // pcie_wr to WC space
    memcpy(nic_q_addr, data, 64);

    // write barrier to flush
    asm("sfence": : :"memory");
}
```

No DMA for minimum size packets

Receive example

```
while(!cache_line_data_valid(cl_addr)){
    asm("clflush %0":"+m" cl_addr));
    // issue a cache miss to the NIC
    pcie_read(cl_addr);
}
if(opt_60B_rx(cl_addr)){
    // entire packet is in cl_addr
    process_60b(cl_addr);
}
else{
    // packet was DMAed to host memory
    process_pkt(cl_addr);
}
```

Summary

- 60B pkt TX+RX: PCIe_wr + NIC_logic + MAC + PHY + cable + PHY + MAC + NIC_logic + PCIe_reply
  1989ns = 484ns + 150ns + 202ns + 135ns + 15ns + 135ns + 438ns + 150ns + 280ns
- 64B pkt TX+RX: PCIe_wr + PCIE_rd + NIC_logic + MAC + PHY + cable + PHY + MAC + NIC_logic + PCIe_write + cache_miss(2x)
  3059ns = 484ns + 850ns + 260ns + 202ns + 135ns + 15ns + 135ns + 438ns + 160ns + 280ns + 100ns