FPGAs and Networking

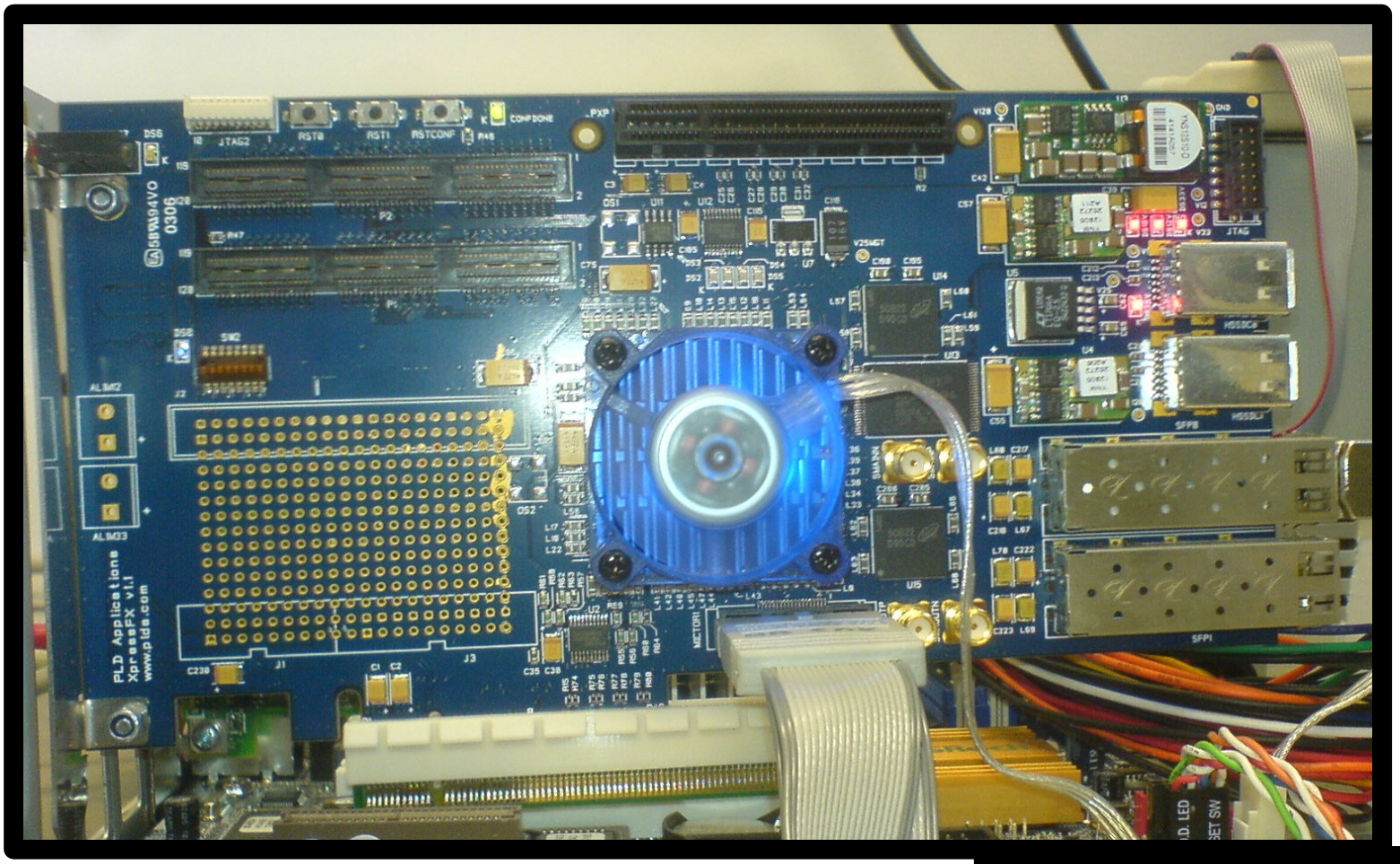
Marc Kelly
& Richard Hughes-Jones

University of Manchester

Overview of Work

- Looking into the usage of FPGA's to directly connect to Ethernet for DAQ readout purposes.
- Testing both 1 and 10 Gig systems.
- Evaluating the new generation of PCIExpress 10Gig Ethernet cards.
- Bringing it all together to form a test system.

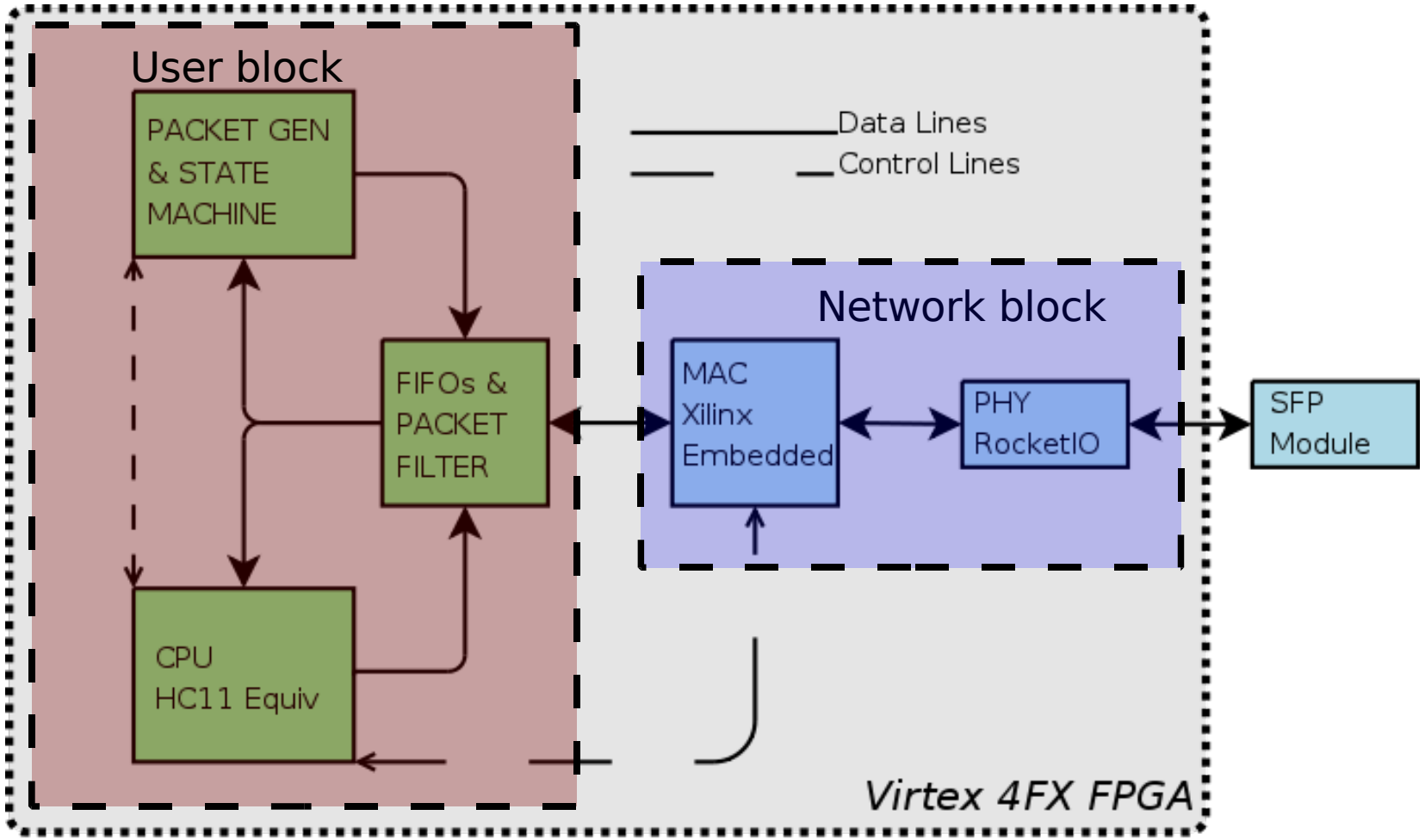
Network Virtex4 Test Board



1 Gig FPGA Work

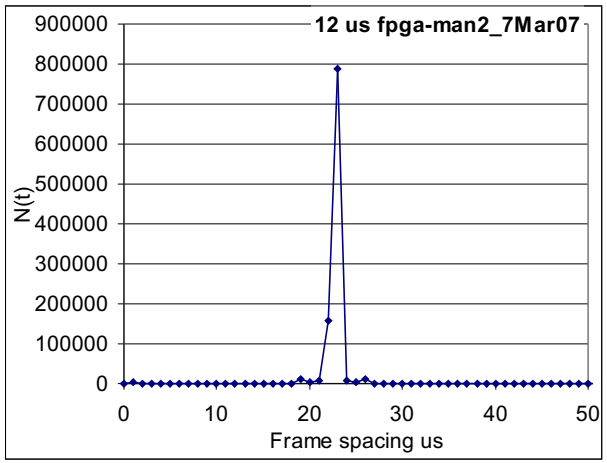
- Implemented a MAC + PHY layer inside Xilinx Virtex4 FPGA.
- Demonstrated working Ethernet between FPGA and remote hosts, however the learning curve is steep, issues with the Xilinx “CoreGen” design.
- Adaptable testing design with Network and “User” FPGA logic separated
- Once working it has proved reliable, the Network code as survived many “re-spins” of the design without failing.

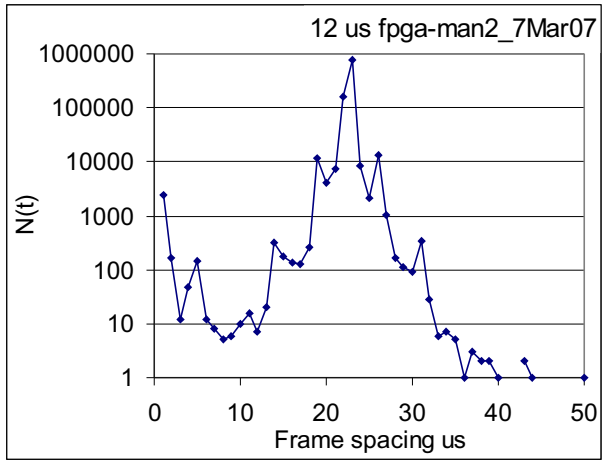
Overview of Design

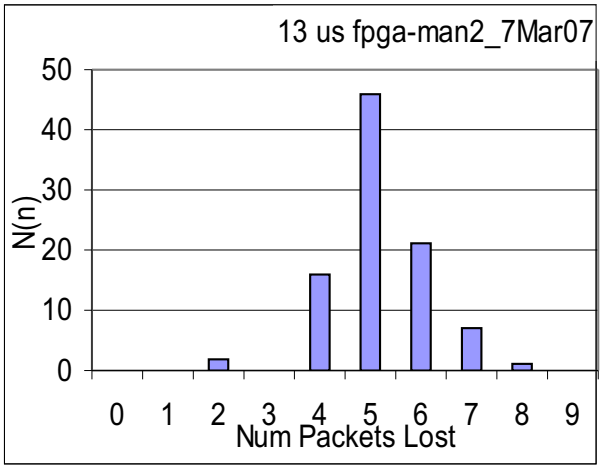


Receiver Frame Jitter and Packet Loss

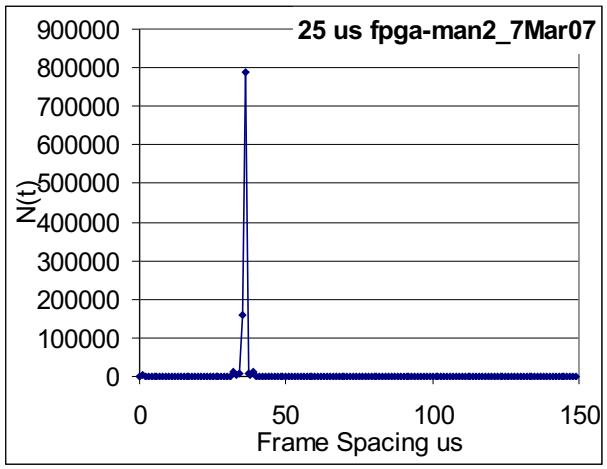
- **12 us (line speed)** **Frame Jitter** **Packet loss**

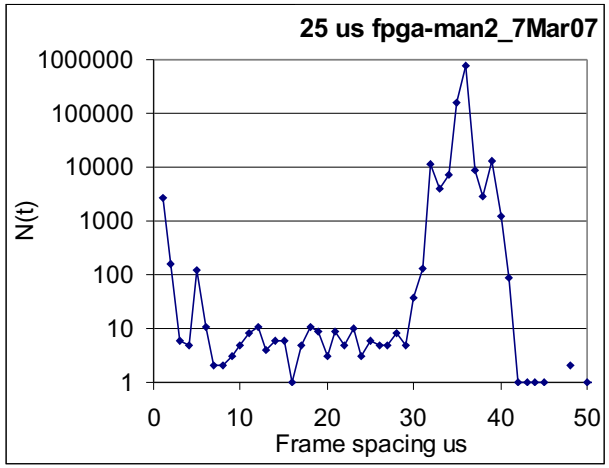


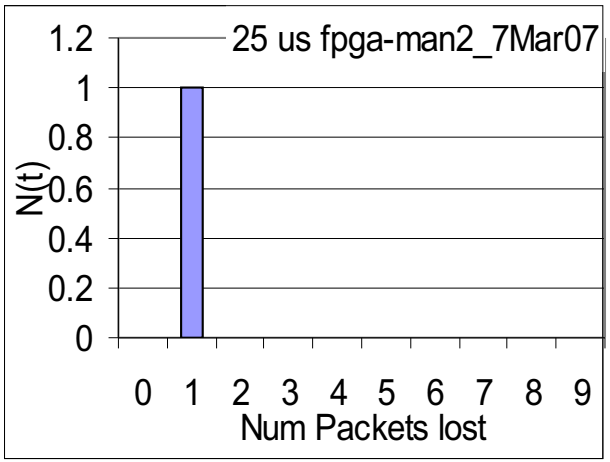




- **25 us frame spacing** **Peak separation 4-5 us no coalescence**







Plots by Richard, stolen from one of his talks.

1 Gig FPGA Work

- FPGAs can drive Ethernet. It is easy once configuration hurdle is passed.
- More testing under way. Request response style operations to pull data out FPGA to simulate an event building scenario.
- Planned Upgrade to 10Gig Ethernet. Do all tests at 10Gig.
- Perform some initial testing to try to determine the stability of the RocketIO TX/RX latency.

10Gig Ethernet Work

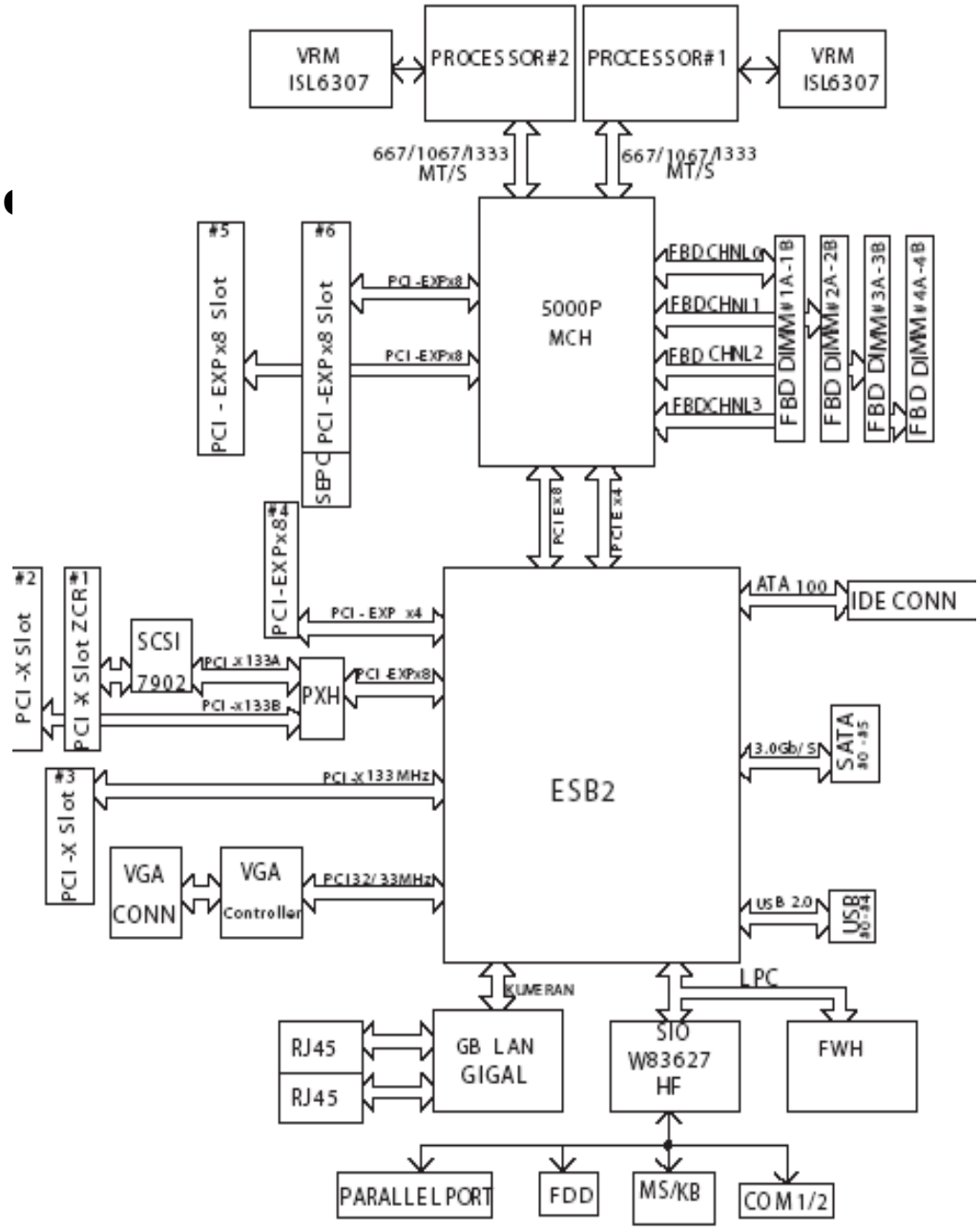
- Richard has nice new 10Gig PCI Express cards made my Myricom. 8xLane design, in theory that has more than enough bandwidth to deal with 10Gig link.
- Using loaned high end servers as host machines.
- Performing standard network testing operations using his normal tools.
- Aim is to determine the suitability of 10Gig systems.

High-end Server PCs

- **Boston/Supermicro X7DBE**
- **Two Dual Core Intel Xeon Woode**
  - 2 GHz
  - Independent 1.33GHz FSBuses
- **530 MHz FD Memory (serial)**
  - Parallel access to 4 banks

- **Chipsets: Intel**
  5000P MCH – PCIe & Memory
  ESB2 – PCI-X GE etc.

- **PCI**
  - 3 8 lane PCIe buses
  - 3* 133 MHz PCI-X
- 2 Gigabit Ethernet
- SATA



10 GigE Back2Back: UDP Throughput

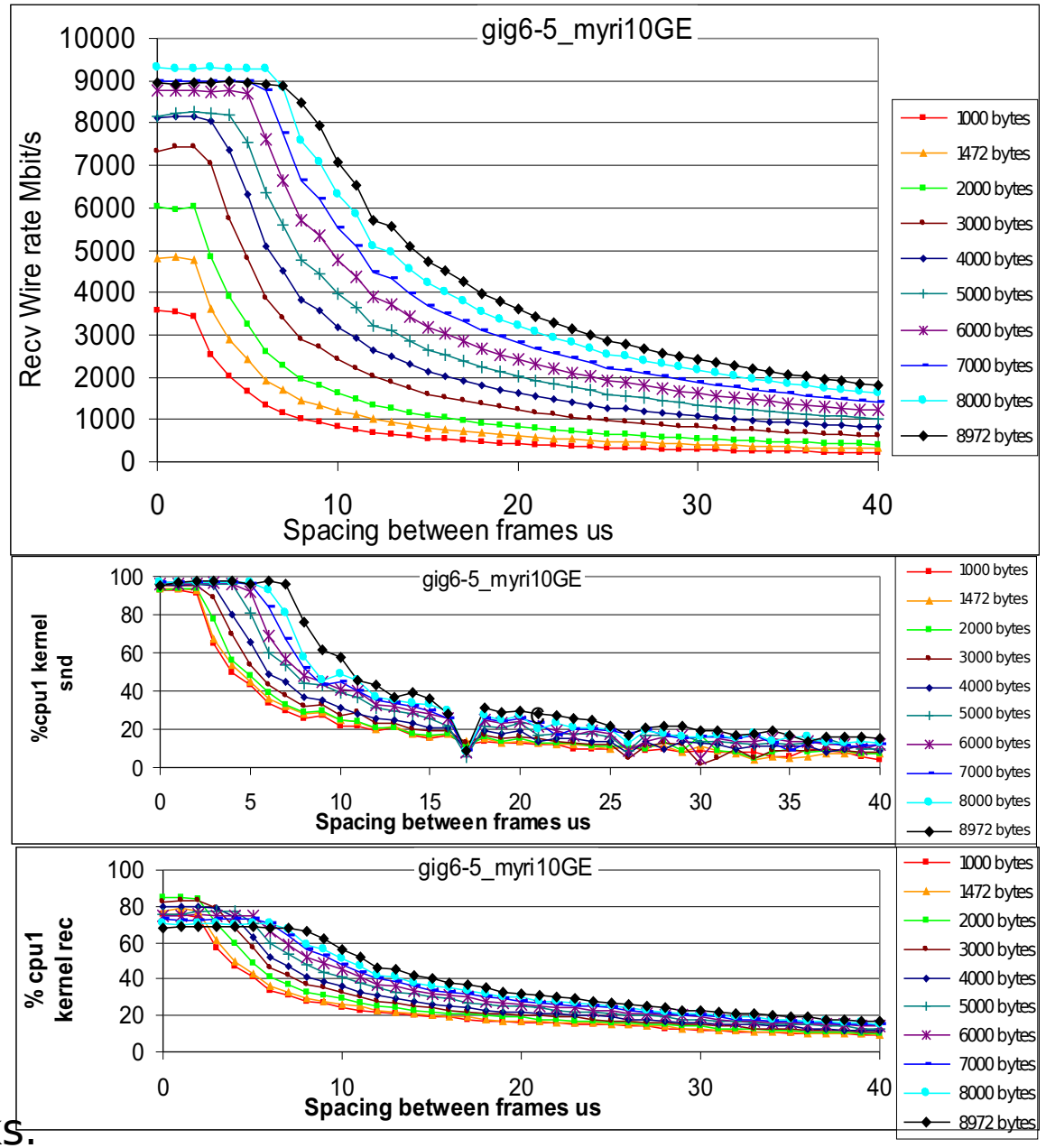
- **Kernel 2.6.20-web100_pktd-plus**
- **Myricom 10G-PCIE-8A-R Fibre**
  - rx-usecs=25
    Coalescence ON
- MTU 9000 bytes
- Max throughput **9.4 Gbit/s**

- Notice rate for 8972 byte packet

- ~0.002% packet loss in 10M packets
  in receiving host

- Sending host, 3 CPUs idle
- For <8 µs packets,
  1 CPU is >90% in kernel mode
  inc ~10% soft int

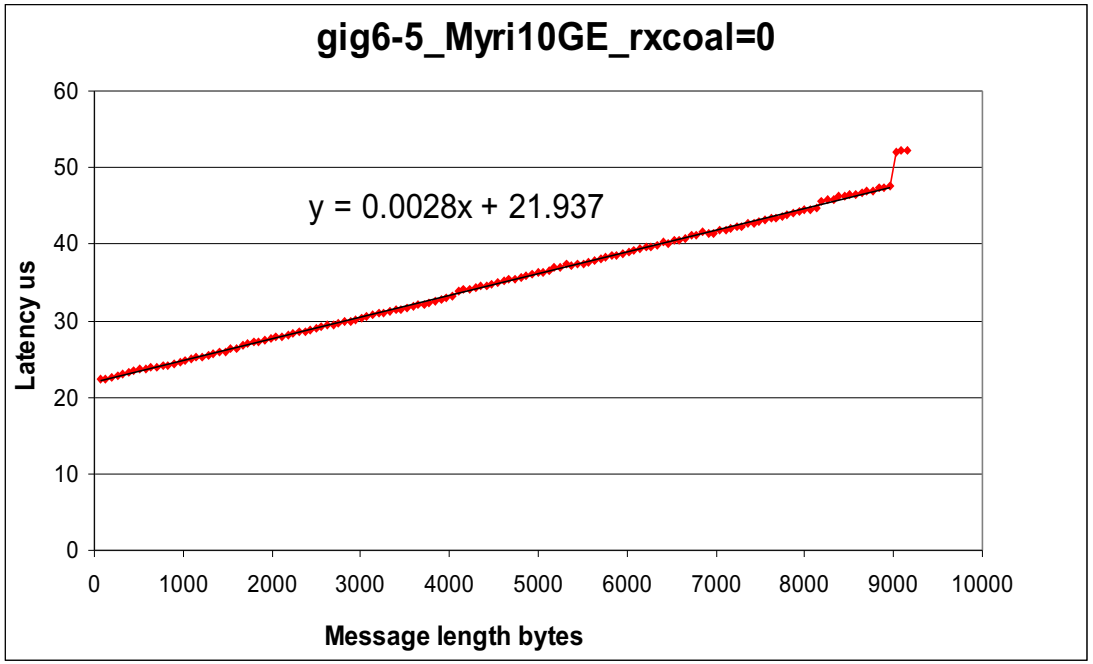
- Receiving host 3 CPUs idle
- For <8 µs packets,
  1 CPU is 70-80% in kernel mode
  inc ~15% soft int



By Richard, stolen from one of his talks.

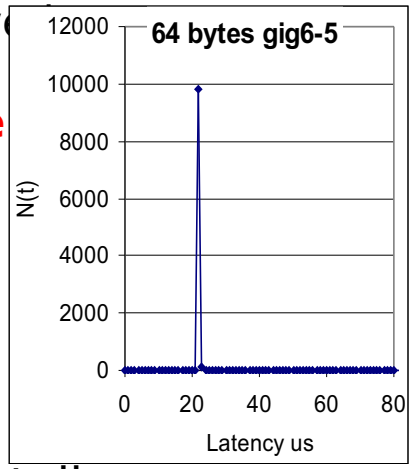
10 GigE Back2Back: UDP Latency

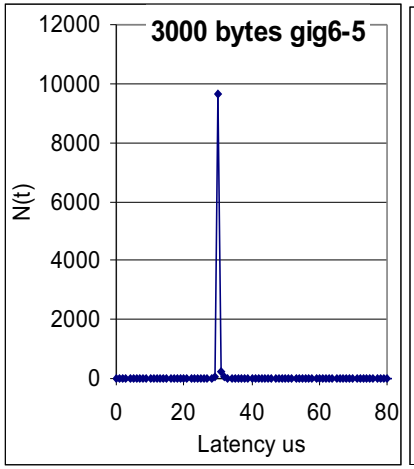
- Motherboard: Supermicro X7DBE
- Chipset: Intel 5000P MCH
- CPU: 2 Dual Intel Xeon 5130 2 GHz with 4096k L2 cache
- Mem bus: 2 independent 1.33 GHz
- PCI-e 8 lane
- **Linux Kernel 2.6.20-web100_pktd-plus**
- **Myricom NIC 10G-PCIE-8A-R Fibre**
- myri10ge v1.2.0 + firmware v1.4.10
  - rx-usecs=0 Coalescence OFF
  - MSI=1
  - Checksums ON
  - tx_boundary=4096
- **MTU 9000 bytes**

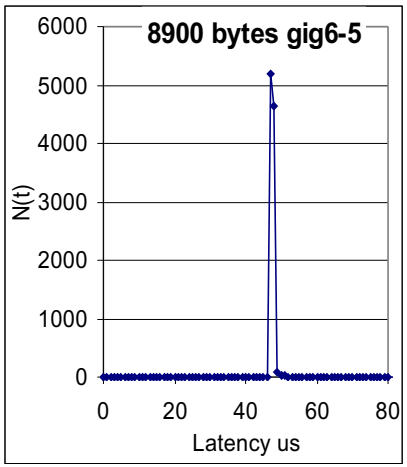


- Latency 22 µs & **very** well behave
- Latency Slope **0.0028 µs/byte**
- B2B Expect: **0.00268 µs/byte**
  - Mem 0.0004
  - PCI-e 0.00054
  - 10GigE 0.0008
  - PCI-e 0.00054
  - Mem 0.0004

- Histogram FWHM ~1-2 us



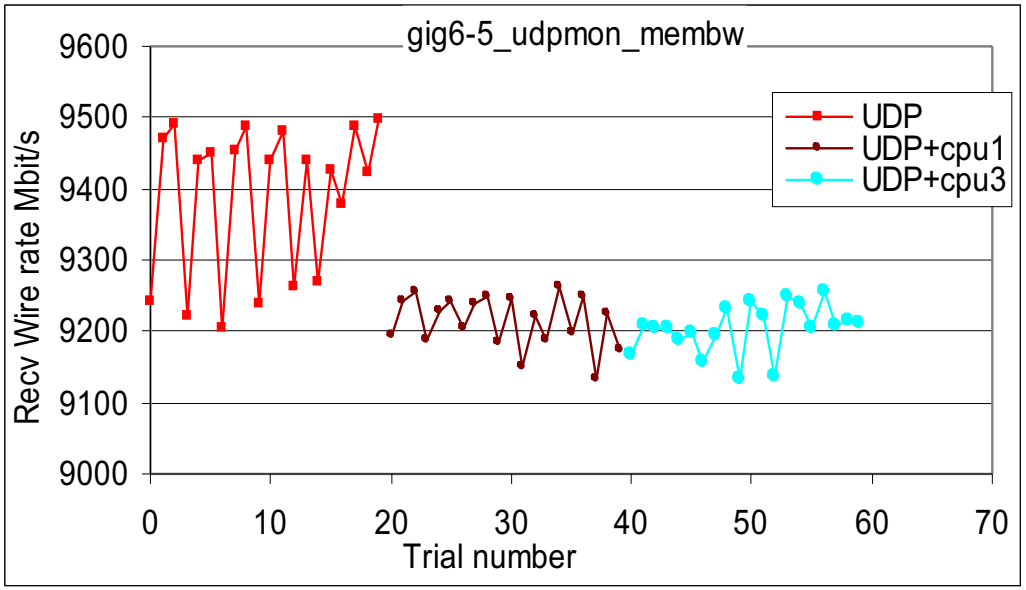


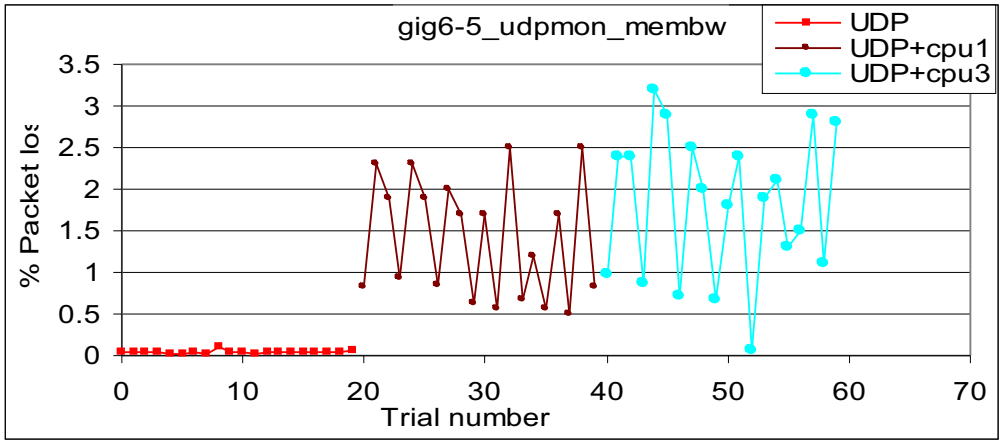


By Richard, stolen from one of his talks.

B2B UDP with memory access

- Send UDP traffic B2B with 10GE
- **On receiver run independent memory write task**
  - L2 Cache 4096 k Byte
  - Write 8000k Byte blocks in loop
  - 100% user mode
- Achievable UDP Throughput
  - **mean 9.39 Gb/s sigma 106**
  - **mean 9.21 Gb/s sigma 37**
  - **mean 9.2 sigma 30**
- Packet loss
  - **mean 0.04%**
  - **mean 1.4 %**
  - **mean 1.8 %**
- CPU load:






```
Cpu0  :  6.0% us, 74.7% sy,  0.0% ni,  0.3% id,  0.0% wa,  1.3% hi, 17.7% si,  0.0% st
Cpu1  :  0.0% us,  0.0% sy,  0.0% ni, 100.0% id,  0.0% wa,  0.0% hi,  0.0% si,  0.0% st
Cpu2  :  0.0% us,  0.0% sy,  0.0% ni, 100.0% id,  0.0% wa,  0.0% hi,  0.0% si,  0.0% st
Cpu3  : 100.0% us,  0.0% sy,  0.0% ni,  0.0% id,  0.0% wa,  0.0% hi,  0.0% si,  0.0% st
```

By Richard, stolen from one of his talks.

10Gig Ethernet Work

- New generation of servers are capable of supporting 10Gig Ethernet, doing real work and NOT being overloaded.
- New generation of Cards are very capable of supporting 10Gig Ethernet.
- Things are however Chipset/Server design dependant. Have to make sure the architecture is sound. High bandwidth, low contention designs needed.
- Need modern host OS, latest drivers etc.

Future Things

- More detailed DAQ development.
- Continued Ethernet work. Finish 1Gig studies and start 10Gig FPGA work. Need to determine if 10Gig upgrade is 100% possible with current board.
- Examining options for the “Link Data Aggregation” board in collaboration with the other DAQ people involved.
- Work on “LDA” link protocol to determine if constant latency system can be bought/built easily.