A Fast Ethernet Tester Using FPGAs and Handel-C

The Large Hadron Collider at CERN
The ATLAS Experiment at CERN
Overview

- ATLAS - the networking requirement
- Switch models and required testing
- Fast Ethernet tester/ROB emulator
- The FE Tester architecture
- Implementation - 36 FPGAs
- Language - Handel-C
  - Problems overcome in version 2.1
  - Issues with version 3.0
ATLAS LVL2 system network

1564 buffers (distributed 1 Mbyte image)

Readout buffers (ROB)

Concentrating switches

Gigabit Ethernet

CENTRAL Gigabit switch

Fast Ethernet

Processing nodes

~500-500MIPS processors analyze data from 5% of buffers

Trigger rate
100 kHz
Commodity Ethernet switches

Most Ethernet switches on the market today have a hierarchical architecture

- ports
- modules
- backplane

Store-and-forward mode of operation

- frame is fully stored in module with an input port
- for inter-module transfer the frame is subsequently stored in a module with an output port
Parameterized model for switch: intra-module communication

Parameters:

- **P1** - Input Buffer Length [ #frames]
- **P2** - Output Buffer Length [ # frames]
- **P5** - Max Intra-module Throughput [MB/s]
- **P8** - Intra-module Transfer Bandwidth [ MB/s]
- **P10** - Intra-module Fixed Overhead [µs] (not shown)
Parameterized model for switch: inter-module communication

Parameters:

P1 - Input Buffer Length [ #frames]
P2 - Output Buffer Length [ # frames]
P3 - Max ToBackplane Throughput [MB/s]
P4 - Max FromBackplane Throughput [MB/s]
P6 - Max Backplane Throughput [MB/s]
P7 - Inter-module Transfer Bandwidth [ MB/s]
P9 - Inter-module Fixed Overhead [µs] (not shown)
Switch parameters from latency and rate measurements

- Measure latency as function of packet size (ping-pong)
- Measure packet rate as function of packet size
- No switch, intra-module, inter-module

<table>
<thead>
<tr>
<th>Setup of the measurement</th>
<th>What is measured</th>
<th>What can be learned from the measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct connect</td>
<td>Latency L</td>
<td>The PCs software overhead</td>
</tr>
<tr>
<td>Intra-module</td>
<td>Latency L</td>
<td>Constant overhead $P10$ and transfer bandwidth $P8$:</td>
</tr>
<tr>
<td>Inter-module</td>
<td>Latency L</td>
<td>Constant overhead $P9$ and transfer bandwidth $P7$:</td>
</tr>
<tr>
<td>Direct connect</td>
<td>Achieved rate</td>
<td>Maximum throughput the PCs can generate/absorb</td>
</tr>
<tr>
<td>Intra-module</td>
<td>Achieved rate</td>
<td>Maximum intra-module throughput: $P5$</td>
</tr>
<tr>
<td>Inter-module</td>
<td>Achieved rate</td>
<td>Maximum inter-module throughput: $P3$, $P4$, $P6$</td>
</tr>
</tbody>
</table>
Limitations of PC-based measurements

- PC with standard software unable to drive switches at line speed
  - 100% for FE only for packet size > 500 Byte [half duplex]
  - Parameterization of a single switch needs to be tested over the entire operational envelope

- Number of nodes limited to 40 in present tests
  - Scaling up to ~2000 nodes is not credible
  - Model needs to be verified on a larger test bed

Conclusion:
- Dedicated traffic generators needed for FE in order to fully characterize switches to be used in ATLAS HLT
- Need larger test bed for greater confidence in model applied to full ATLAS HLT system
Fast Ethernet tester

- Custom board, built at CERN
- Programmable network traffic to characterize network switches
- IP, Ethernet, including QoS
- Why build it instead of buying?
  - Economics
  - Reprogrammable for ATLAS purposes
FE tester: capabilities & status

- 32 ports full-duplex fast Ethernet
- Parallel port connection to host
- MAC function: FPGA
- All FPGAs programmed in Handel-C
  - Version 2.1
- Single 25 MHz hardware clock for all ports
- Global clock synchronized among multiple testers
- Output packets generated, time stamped
- Incoming packets time stamped, CRC checked
- Network latencies histogrammed
- Full duplex @ full line speed
Fast Ethernet tester board
CERN Handel-C
32 port Ethernet tester
100Base-T

1.75 million reconfigurable gates
35 Handel-C devices
FE Tester building blocks: the slotted ring

- FE tester backbone interconnects all of the MACs, RxMan, TxMan, and IOMan
- 54 bits wide
- Synchronous daisy chain running at 25MHz
- Each MAC gets a time slot every 6.44 µsec
- Each MAC can load and unload descriptors from the ring during its time slot
- Descriptors: 20 bytes
FE Tester building blocks: MAC

- Altera FLEX 10KS50 (256 pin package)

- Tx processor
  - Programmable
  - Runs code which allows flexible processing of Tx descriptors

- Rx processor
  - Hard-coded to extract information from incoming packets

- Handel-C (version 2.1)
  - Multiple clock domains not supported
  - Required some bits of VHDL
  - Gluing together VHDL and Handel-C required major work arounds
  - Lacked support for dual-port RAM
FE Tester building blocks: **TxMan**

- Altera FLEX 10KE50 (484 pin package)
- 1 Mword (36 bit) directly attached SRAM
- **TxMan** generates transmit descriptors for each MAC using descriptors stored in SRAM
- Descriptors contain enough information for the MAC Tx processor to generate packets, e.g.,
  - Index of dest port
  - Packet length
  - Packet type
  - VLAN tag
- **Descriptor format is flexible**
  - Only constraint is 20-byte limit
- **6500 descriptors per FE port**
  - Systematic or randomized
FE Tester building blocks: **RxMan**

- Altera FLEX 10KE50 (484 pin package)
- 1 Mword (36 bit) directly attached SRAM
- RxMan receives and processes descriptors for incoming packets generated by the MACs
- Extracts information from the descriptor, e.g.,
  - Source address
  - Packet type
  - Latency
  - VLAN tags (priority)
- Updates packet, byte, latency counters
- Updates histograms
- Generate response descriptors
- It has 5 clocks to process each descriptor
  - 2 reads + 2 writes to external RAM
  - 3 reads + 3 writes to internal RAM
FE Tester building blocks: IoMan

- Altera FLEX 10KS50 (256 pin package)
- IoMan communicates with the host via a fast (IEEE1284) parallel port
- Relays all commands to TxMan, RxMan, MACs
- Relays results from TxMan, RxMan, MACs
- Control and status of ethernet physical I/O ports
- Global clock synchronization
- Configures other FPGAs via serial bit stream
Handel-C compilation: 3 components

• Handel-C compiler
  ▪ Windows NT host

• Perl scripts (Linux host)
  ▪ Glue together VHDL and Handel-C
  ▪ Implement on-chip dual-ported RAM not supported by Handel-C version 2.1
  ▪ Rename compiler-generated symbols to aid diagnostics

• Altera MaxPlus tools
  ▪ Sun Solaris host
FE tester - downloading code

• IoMan
  • JTAG via Byte Blaster cable
    • during development
  • IoMan code is very stable
    • Autoload from serial EEPROM

• Other FPGAs
  • Bitstreams downloaded to IoMan, relayed to target devices using serial programming
The user interface
Latency measurement for two priorities

Latency vs offered throughput

Throughput (% line speed)

- low priority latency
- high priority latency
FE tester as ROB emulator

- Accepts ROB DATA REQUESTs

- Generates ROB DATA REPLYs
  - Programmable latency in ROB
  - Response size depends on ROI size and detector type
  - Response contents not meaningful

- Measure latencies and queue depths
Large scale test bed using ROB emulator

- Use 8 of these boards [11K CHF each] to provide 256 emulated ROB ports

- Add:
  - 64 PCs - supervisors + farm nodes
  - Switch fabric

- We have a ~15% scale test bed of Atlas LVL2
  - Previous test bed: order of magnitude smaller
Test bed capabilities

• Supervisor (PC) generates ROI patterns using simulated events
• Measure request-response latency
• Measure queue dwell times in ROB
  ▪ Timestamp on request arrival and response generation
• Measure queue occupancy distribution
  ▪ Not possible with software emulation
• Measure frequency of broken messages under realistic conditions in real switches
Problems with Handel-C version 2.1

• Multiple clock domains not supported
• Dual-ported RAM not supported
• Integration of VHDL code not supported
• Messages from FPGA vendor tools show meaningless compiler-generated names
### Outstanding issues with Handel-C version 3.0

- Both compilers were fed the same source code
- Achieved clock speeds not yet usable (25 MHz required)

<table>
<thead>
<tr>
<th>Chip</th>
<th>Handel-C 2.1 clock speed</th>
<th>Handel-C 3 clock speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOMAN</td>
<td>42.55 MHz</td>
<td>32.46 MHz</td>
</tr>
<tr>
<td>RXMAN</td>
<td>41.49 MHz</td>
<td>23.14 MHz</td>
</tr>
<tr>
<td>TXMAN</td>
<td>31.15 MHz</td>
<td>24.50 MHz</td>
</tr>
<tr>
<td>MAC</td>
<td>27.93 MHz</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- Existing MAC Tx processor code does not compile
  - New compiler creates logic cycles not created by version 2.1
Summary

• Constructed 32-port FE tester
• 36 FPGAs programmed almost entirely in Handel-C
  • 3 active Handel-C users!
• Project is nearing completion
• Need to resolve outstanding Handel-C version 3 difficulties