Common Gigabit Ethernet interfaces for HLT and L1-trigger links of LHCb

On behalf of the LHCb collaboration

Hans Muller, Francois Bal, Angel Guirao
CERN ED group
Gigabit Ethernet Physical Layer (PHY)

4 lines @ 250 Mbit/s each Direction
1000BASE-T link = 4 X bidir twisted pair

- Marvell-Alaska PHY chip is better than 802.3ab
- up to 180 m UTP with BER= 10E-10

IEEE 802.3ab : up to 100 meter CAT 5 cable
4 pairs Unshielded Twisted Pair (UTP)
4 duplex Rx/Tx lines @ clock 125 MHz
4 lines @ 250 Mb/s, 5 level PAM coding

Phy chips perform auto-negotiation
100/1000BASE-T
LHCb’s GBE mezzanines

Reference design described here
2-Channel Rx Tx

Under consideration:
4-Channel Rx Tx
Geometrical Outlines

VME airflow cooling:
8 Watt / mezzanine

High speed Point-Point signalling
Between motherboard and mezzanine:

Component heights
SPI-3 100 MHz Connector

- **Mezzanine Connector**
- **SAMTEC:** QSS-075-01-F-D-A-xx
- 3 banks 25 positions (3 * 50 pins)
- 0.635 mm spacing
- mates with QTS-075-03-F-D-A-XX
- 11 mm stack-height
- Surface mount
- 25 pin groups centered by bus-bar
- Each pin 50 Ohm to bus-bar
Dual MAC chip

Figure 2 - PM3386 Dual Gigabit Ethernet to POS-PHY Level 3

- 4 Tx handsh
- 4 pause
- TDA[31:0] (50)
- 10 Ctrl Tx

- 2 Rx handsh
- RDA[31:0] (50)

- 8 Rx

- 5 JTAG
- 2 (5 Volt)
- 34 generic

16-bit ctrl. bus (POS-PHY-Level3)
SPI-3 industry bus

generic:

SPI-3 bus (8, 32)

media

MAC

PHY

Link Device

switch

LHCb:

cat5e

SPI-3 bus 32 bit

2*MAC

2*1 GBps

L1 buf

TX

RX

FPGA

SPI-3 bus between MAC chip and FPGA of Tell-1

LHCb: send data encapsulated in IPv4 frames

Terminology:

based on “POS-PHY-Level3” of SATURN Group

SPI-3 is networking industry standard for OC-48

48 * OC-1 (51.84 Mbit/s) = 2.488 Gbit/s

SPI-3 Tx protocol (simplified):

1.) TPA signals free space in FIFO

2.) TENB: select PHY addr (HLT or L1) and data packet

FPGA monitors TPA fill state during transmission

FPGA can pause by lowering TENB

FPGA polls the PHY available status

PHY status polling during transfer (in packet mode)

LHCb: send data encapsulated in IPv4 frames

Screenshot on GBE card
Example VELO

VELO trigger input to Level-1 latency buffer: 40 byte @ 1 MHz

Level-1 buffer module (Tell-1):
- Preprocessing, event-buffering,
- Prompt trigger extraction (L1T stream)
- Level-1 decision: (HLT stream)
- Multi-Event-Packing in FPGA (MEP>IPv4)
- Transmission to GBE card via SPI-3

2-channel GBE mezzanine:
Transmit two directed data streams: HLT and L1T as IPv4 packets to GBE switched network

H1T stream: 76 sources per link ~10 MB/s

L1 average stream: 76 sources per links @ 40 MB/s each if sent as MEP

Level-0 trigger out: max. 1.1 MHz

Level-1 trigger out: max 40 kHz
**Multi Event Packets (MEP)**

- **L1T output raw:** 36 byte
  - 40 byte @ 1 MHz
  - 1 MHz Ethernet rate too high
  - Pack up to 32 L1 fragments into 1 MEP packet
  - **L1T:** 34 kHz MEP\(_{32}\) packets

1. Hardware driver in FPGA makes MEPs
2. MEP are formatted like IPv4 Ethernet
3. Send to MAC chip via SPI-3

**Diagram:**
- **FPGA** (Tell-1)
- **MAC** (GBE card)
- **SPI-3 bus**
LHCb data format over IPv4

8 6 6 2 46....1500 4 bytes

Pre-amble  Dest. address  Source address  Len/Type protocol  Data (+ protocol)  FCS

Ethernet Header  data payload

Len/Type field > 0x600  Type o. protocol

To be created in FPGA
And transmitted to GBE
FPGA driver coding

Destination address assignment:

Destination address from TTC long broadcast (channel B):

TTC

10 = destination address

Broadcast

15

reserved

10 bit SFC address

HLT=1 / L1=0

Flush MEP

Ethernet dest. address

10 bit SFC address

FPGA:

IP dest address

10 bit SFC address

destinations:

max 1024 SFC’s

10 bit via TTC from

Readout Supervisor

Data (32 bit up-counter)
Photo 2-channel GBE card

- 2*RJ45
- dual magnetics
- dual PHY controller
- dual MAC controller
- LHCb connector
- 100 MHz
- SPI-3 bus
- Egress
- Ingress
- MAC/PHY control registers
- 3.3 Volt bus bar
- CERN Reference
- 3 blocks 50 pin:
- GND bus-bars
- optical transceiver emplacement (option)
- * 1 Gbps bidirectional
- (or 2nd Egress)

1 channel
100Base-T

3 blocks 50 pin:
- 2 * 1 Gbps bidirectional

Manufacturing Date (China)

- Photo 2-channel GBE card
FPGA driver on Aroc PCI card

(Aroc: see poster)

Aroc card test platform

PCI bus

2ch- GBE mezzanine
Aroc test system

Aroc PCI card

GBE card
Test environment

Mac
Register
Initialization

Network
Analyser
Software

Labview

PC-1
Aroc dll
Aroc PCI card
SPI-3 bus

PC-2
1000Base-T

FPGA
MAC
Dual PHY
Dual Magnetics
RJ45
RJ45

hardware
IP driver

Aroc.dll
hardware
IP drive

Labview

Network
Analyser
Software
First IPv4 packets

Network sniffer on destination PC

FPGA encoding

<table>
<thead>
<tr>
<th>No.</th>
<th>Time</th>
<th>Source</th>
<th>Destination</th>
<th>Protocol</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>1016</td>
<td>106.666508</td>
<td>200.0.0.1</td>
<td>200.0.137.138</td>
<td>IP</td>
<td>IPv6 hop-by-hop option</td>
</tr>
<tr>
<td>1017</td>
<td>106.666512</td>
<td>200.0.0.1</td>
<td>200.0.137.138</td>
<td>IP</td>
<td>IPv6 hop-by-hop option</td>
</tr>
<tr>
<td>1018</td>
<td>106.666515</td>
<td>200.0.0.1</td>
<td>200.0.137.138</td>
<td>IP</td>
<td>IPv6 hop-by-hop option</td>
</tr>
<tr>
<td>1019</td>
<td>106.666519</td>
<td>200.0.0.1</td>
<td>200.0.137.138</td>
<td>IP</td>
<td>IPv6 hop-by-hop option</td>
</tr>
<tr>
<td>1020</td>
<td>106.666523</td>
<td>200.0.0.1</td>
<td>200.0.137.138</td>
<td>IP</td>
<td>IPv6 hop-by-hop option</td>
</tr>
</tbody>
</table>

**Example of frame**

| MAC DA | 00 07 E9 17 | 83 00 00 00 |
| MAC SA | 00 00 00 00 | 00 00 00 00 |

<table>
<thead>
<tr>
<th>Frame 10701</th>
<th>100 bytes on wire, 100 bytes captured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethernet II, Src: 00:00:00:00:00:00, Dest: 00:07:E9:17:83:00</td>
<td></td>
</tr>
<tr>
<td>Destination: 00:07:E9:17:83:00 (00:07:E9:17:83:00)</td>
<td></td>
</tr>
<tr>
<td>Source: 00:00:00:00:00:00 (00:00:00:00:00:00)</td>
<td></td>
</tr>
<tr>
<td>Type: IP (0x0800)</td>
<td></td>
</tr>
<tr>
<td>Version: 4</td>
<td></td>
</tr>
<tr>
<td>Header length: 20 bytes</td>
<td></td>
</tr>
<tr>
<td>Differentiated Services Field: 0x00 (DSCP 0x00: Default: ECN: 0x00)</td>
<td></td>
</tr>
<tr>
<td>0x00 00.. = differentiated services codepoint: default (0x00)</td>
<td></td>
</tr>
<tr>
<td>..... 0. = ECN-Capable Transport (ECT): 0</td>
<td></td>
</tr>
<tr>
<td>..... 0. = ECN-CE: 0</td>
<td></td>
</tr>
<tr>
<td>Total Length: 86</td>
<td></td>
</tr>
<tr>
<td>Identification: 0x0000 (0)</td>
<td></td>
</tr>
<tr>
<td>Flags: 0x04</td>
<td></td>
</tr>
<tr>
<td>... = Don't fragment: set</td>
<td></td>
</tr>
<tr>
<td>... = More fragments: Not set</td>
<td></td>
</tr>
<tr>
<td>Fragment offset: 0</td>
<td></td>
</tr>
<tr>
<td>Time to live: 0</td>
<td></td>
</tr>
<tr>
<td>Protocol: IPv6 hop-by-hop option (0x00)</td>
<td></td>
</tr>
<tr>
<td>Header checksum: 0x61c (correct)</td>
<td></td>
</tr>
<tr>
<td>Source: 200.0.0.1 (200.0.0.1)</td>
<td></td>
</tr>
<tr>
<td>Destination: 200.0.137.138 (200.0.137.138)</td>
<td></td>
</tr>
<tr>
<td>Data (86 bytes)</td>
<td></td>
</tr>
</tbody>
</table>

| 0000 | 00 07 E9 17 83 00 |
| 0010 | 00 07 E9 17 83 00 |
| 0020 | 00 07 E9 17 83 00 |
| 0030 | 00 07 E9 17 83 00 |
| 0040 | 00 07 E9 17 83 00 |
| 0050 | 00 07 E9 17 83 00 |
| 0060 | 00 07 E9 17 83 00 |

| 0000 | 00 07 E9 17 83 00 |
| 0010 | 00 07 E9 17 83 00 |
| 0020 | 00 07 E9 17 83 00 |
| 0030 | 00 07 E9 17 83 00 |
| 0040 | 00 07 E9 17 83 00 |
| 0050 | 00 07 E9 17 83 00 |
| 0060 | 00 07 E9 17 83 00 |
Summary

- 2-channel GBE card reference design successful
- Revision for final LHCb 2-channel GBE card in coming weeks
- PCB reproduction by Tsinghua Univ. Collaboration (Beijing) very good result
- Aroc-PCI same FPGA+GBE plug as LHCb Tell-1 cards (equivalent FPGA developments)
- 2 Aroc cards operational, W2000 PCI driver Aroc.dll
- First GBE tests on Aroc PCI card successful
- MAC register initialization via Labview operational
- First version of FPGA-driver via 100 MHz SPI-3 protocol successful
- LHCb-like IPv4 packest successfully generated in FPGA
- Next:
  - Quad GBE card design ( started )
  - Generate MEPs in Aroc’s memory and transmit at 40 kHz via FPGA driver
  - Copper and optical BER tests
  - preparing for full production GBE cards
Addena

New Intel chip for design of quad GBE card

Intel® IIXF1104 4-port 10/100/1000 Mbps Ethernet Media Access Controller

With System Packet Interface Level 3 (SPI-3)

Product Highlights
- Configurable on a per-port basis for 10/100/1000 Mbps Ethernet
- Includes flexible MAC, and SerDes and Copper PHY interfaces for low-cost connectivity
- Features 32 Kbytes of receive buffer and 10 Kbytes of transmit buffer for lossless flow control of Jumbo Frames
- Flexible system interface supports 32 bit SPI-3 or 4x8 bit SPI-3 up to 133 MHz

Product Overview
The explosion of Internet Protocol (IP) traffic and the industry-standard System Packet Interface (SPI-3) providing a high-performance interface to the Intel® IIXF2400 network processor or to a switching ASIC.

The industry-standard Gigabit Media Independent Interface (GMII) and Reduced Gigabit Media Independent Interface (RGMII) provide the PHY interface for copper connectivity. An integrated serializer/deserializer (SerDes) provides a PHY interface for direct connection to optical modules, helping reduce board real estate and system cost.

The integrated LED controller, internal management data input/output (MDIO) interface with auto-scan capability, high port count, and
Labview test environment

R/w registers of PMC-Sierra chip
Ingress and Egress FiFos
Configuration PHY A/B

http://cern.ch/ep-div-ed/Documents/Mini-these-Antonis.pdf
“X-ray” of 2-channel PCB
First PCBs made in China

LHCb collaboration with Tsinghua University Beijing
Aroc with mezzanine card