LHCb Upgrade Architecture Review
Back End

J.P. Cachemiche, P.Y. Duval, F. Hachon
R. Le Gac, F. Réthoré
Centre de Physique des Particules de Marseille
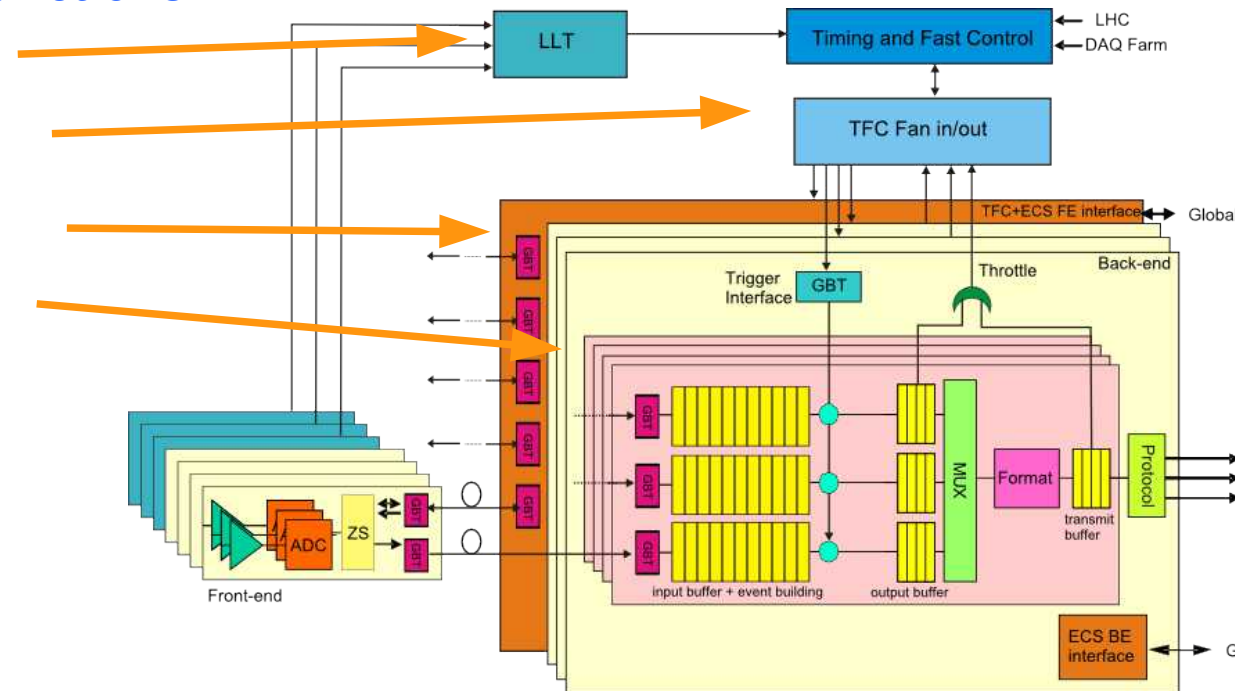
Outline

- **Aim(s) of the board**

- **Board architecture**

- **Data paths: Acquisition, TFC, ECS, IPMI, ...**

- **Crate**

- **Feasibility**

- **Firmware development**

Aims(s) of the board

A single board to address 4 functions :

- **Low Level trigger** (LLT)
- **Timing and fast Control** + fanout (TFC)
- **TFC+ECS** interface to front-Ends (SOL40)
- **Back-End interface** (TELL40)

A board instanciates a specific function
by programming specifically its FPGAs and reorganizing its data paths.

Board architecture

ATCA standard

The LHCb collaboration decided to implement the readout on ATCA

- **Many advantages:**
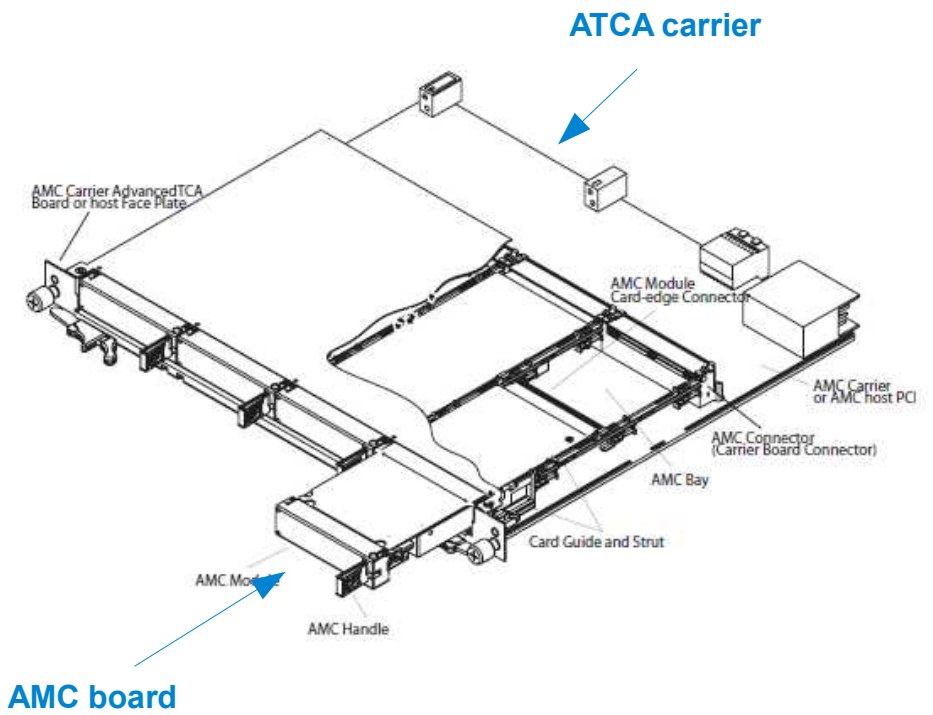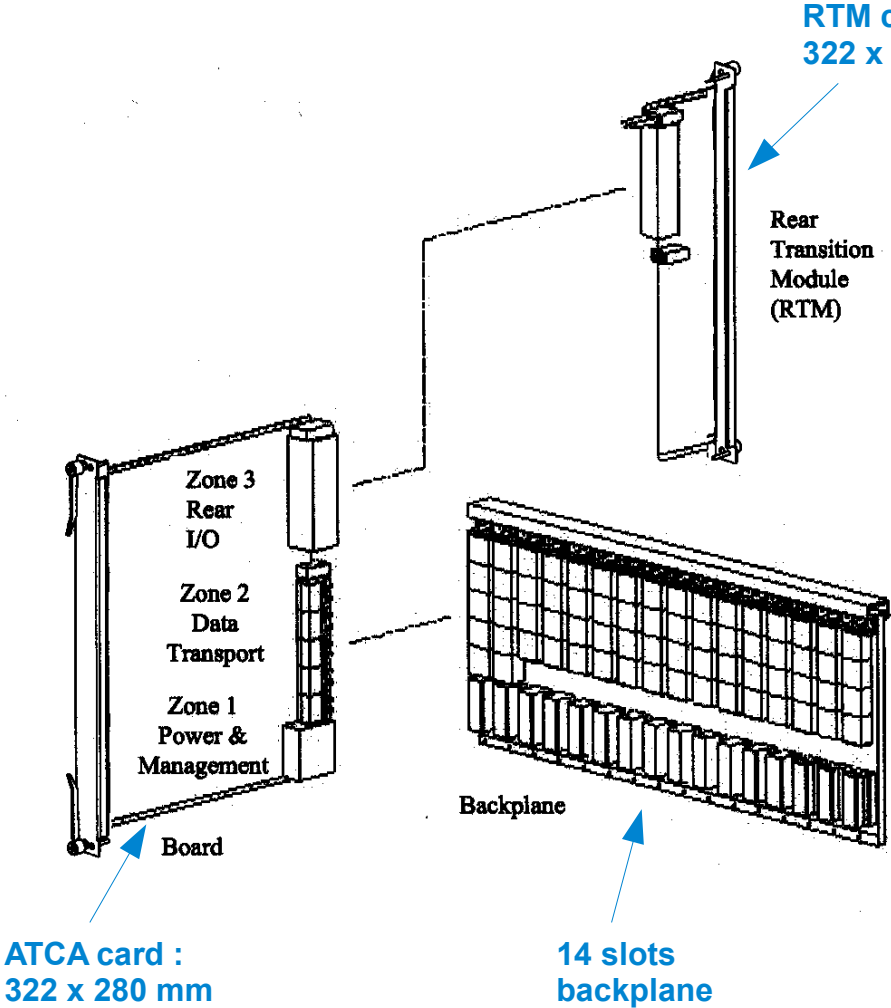 - Robust and well defined mechanics
  - Adapted to recent components
    - *Form factor lets more room for heatsinks*
    - *Power supply dimensionned for high speed components*
    - *Powerfull cooling*
  - Standard backplane
    - *Topology based on serial links*
  - Standard mezzanines
  - Costs comparable to VME
  - Elaborate health monitoring system (IPMI)

- **Difficulties :**
  - IPMI implementation quite complicate

→ Review of feasibility at the end of full scale prototype

ATCA objects

RTM card :
322 x 92/68 mm

Rear
Transition
Module
(RTM)

Zone 3
Rear
I/O

Zone 2
Data
Transport

Zone 1
Power &
Management

Board

Backplane

ATCA card :
322 x 280 mm

14 slots
backplane

ATCA carrier

AMC Carrier AdvancedTCA
Board or host Face Plate

AMC Module
Card-edge Connector

AMC Carrier
or AMC host PCI

AMC Connector
(Carrier Board Connector)

AMC Bay

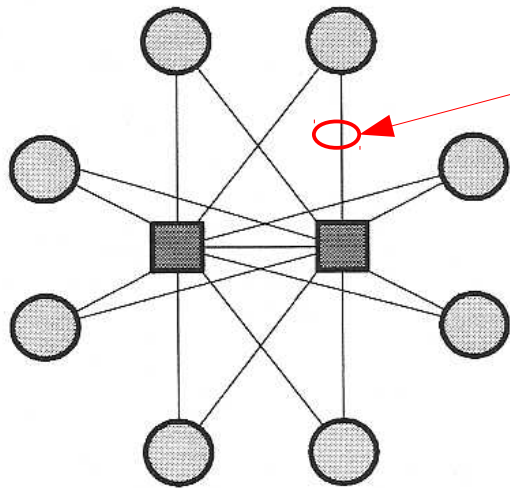Card Guide and Strut

AMC Module

AMC Handle

AMC board

ATCA topologies

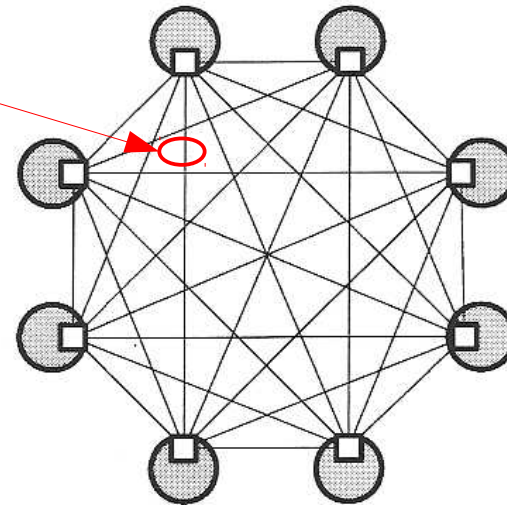## Off the shelf backplanes

- **Exist in several topologies**
  - Dual star
  - Full mesh
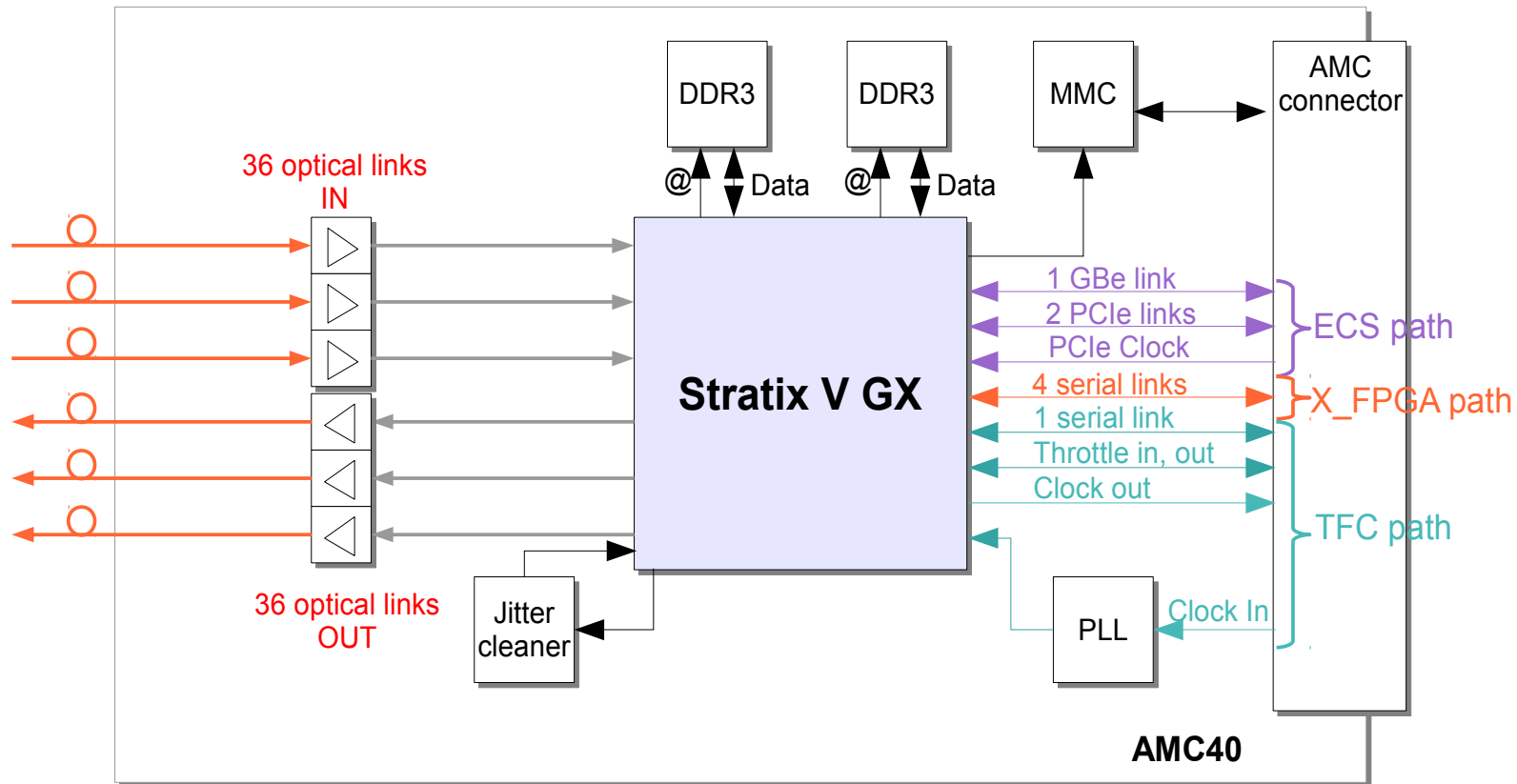
One lane
=
4 bidirectional
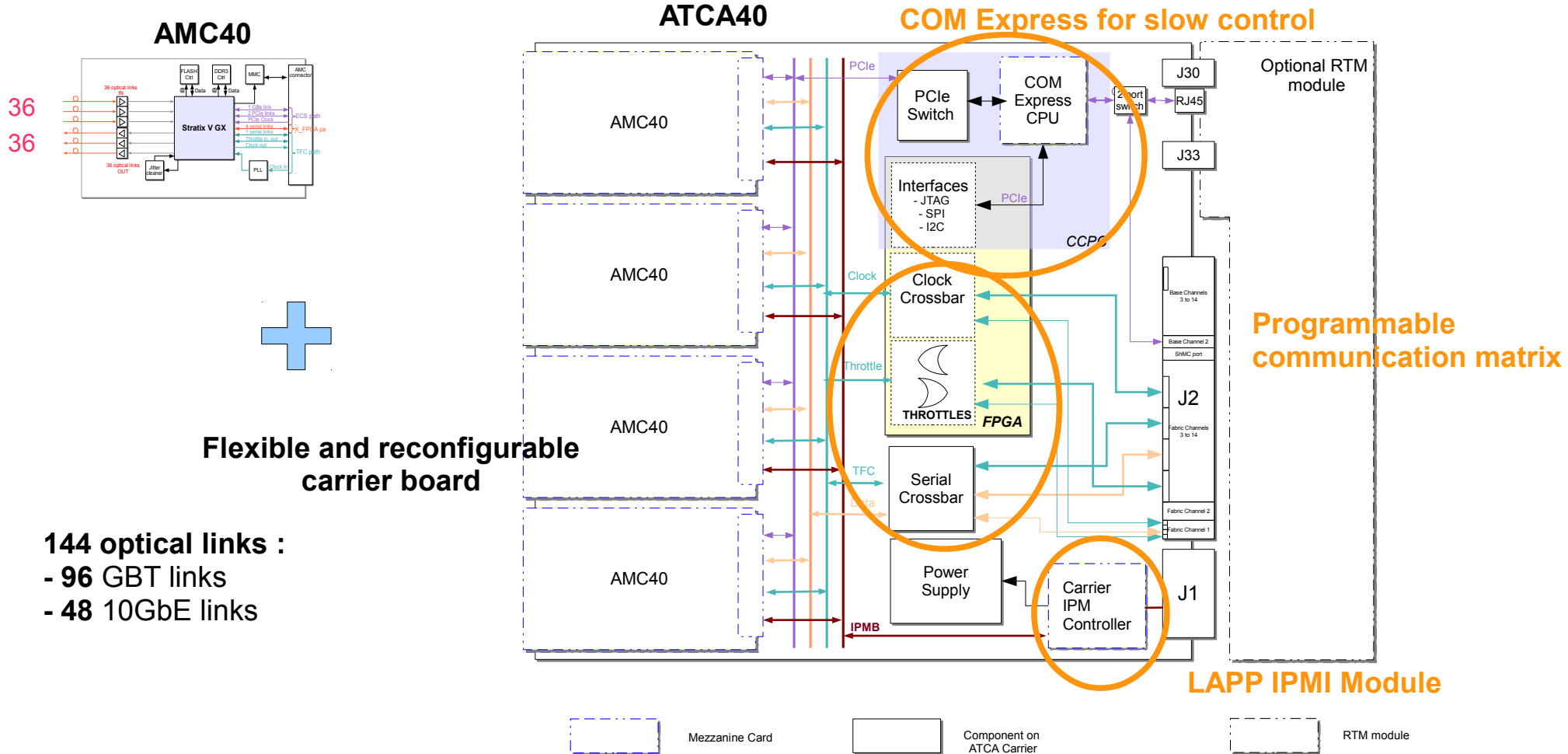serial links

Dual star

Full mesh

Generic optical mezzanine: AMC40

36 bidirectional optical links at up to 10 Gbits/s
622 kLE FPGA Stratix V GX: 5SGXEA7N2F45C3N

AMC40 first prototype

AMC40
14 layers PCB
1 Stratix V GX
36 optical inputs and
36 optical outputs at up to 10 Gbits/s
Slow control through PCIe
Estimated power consumption: 50 W

Generic readout board : ATCA40

AMC40

36
36

+

Flexible and reconfigurable carrier board

144 optical links :
- **96** GBT links
- **48** 10GbE links

ATCA40

COM Express for slow control

Programmable communication matrix

LAPP IPMI Module

Mezzanine Card   Component on ATCA Carrier   RTM module
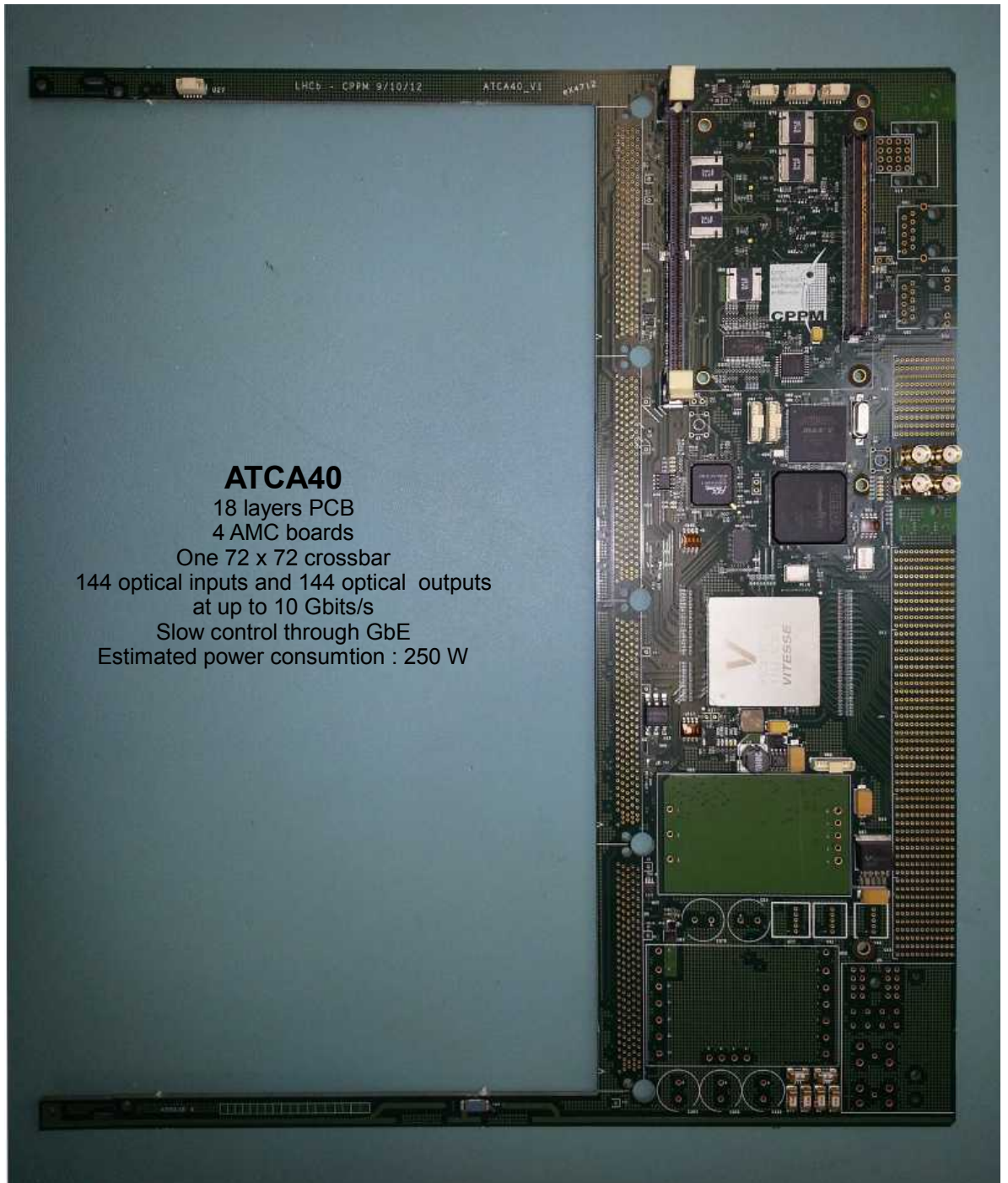
Only firmware and **datapath programming** change
to implement readout, time distribution, slow control or trigger interface

ATCA40
First prototype

ATCA40
18 layers PCB
4 AMC boards
One 72 x 72 crossbar
144 optical inputs and 144 optical outputs
at up to 10 Gbits/s
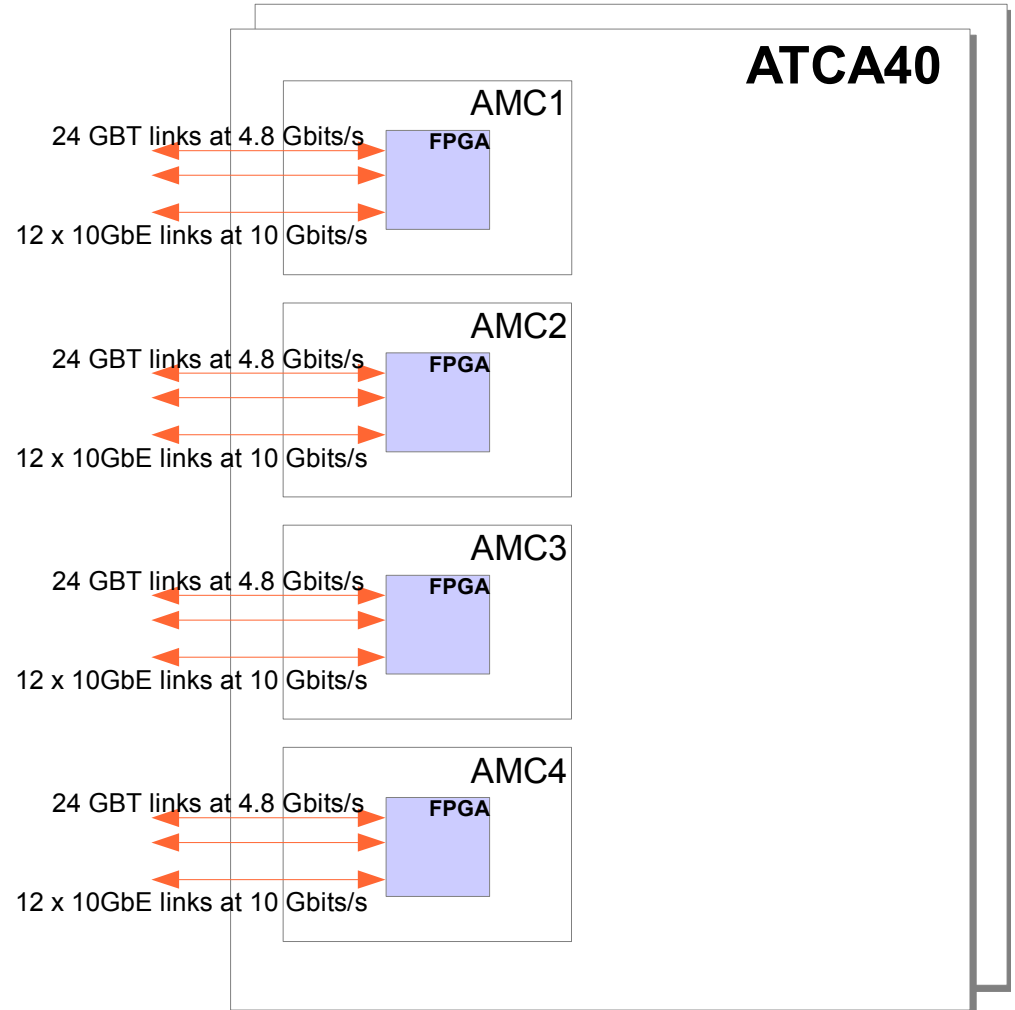Slow control through GbE
Estimated power consumtion : 250 W

Data paths

External Data paths

Optical links on face plate

- 144 bidir links

    - 4 x 24 = 96 bidir links at 4.8 Gbits/s

    - 4 x 12 = 48 bidir links at 10 Gbits/s

ATCA40

AMC1
FPGA
24 GBT links at 4.8 Gbits/s
12 x 10GbE links at 10 Gbits/s

AMC2
FPGA
24 GBT links at 4.8 Gbits/s
12 x 10GbE links at 10 Gbits/s

AMC3
FPGA
24 GBT links at 4.8 Gbits/s
12 x 10GbE links at 10 Gbits/s

AMC4
FPGA
24 GBT links at 4.8 Gbits/s
12 x 10GbE links at 10 Gbits/s

Internal Data paths

Crossbar switch for high speed serial links
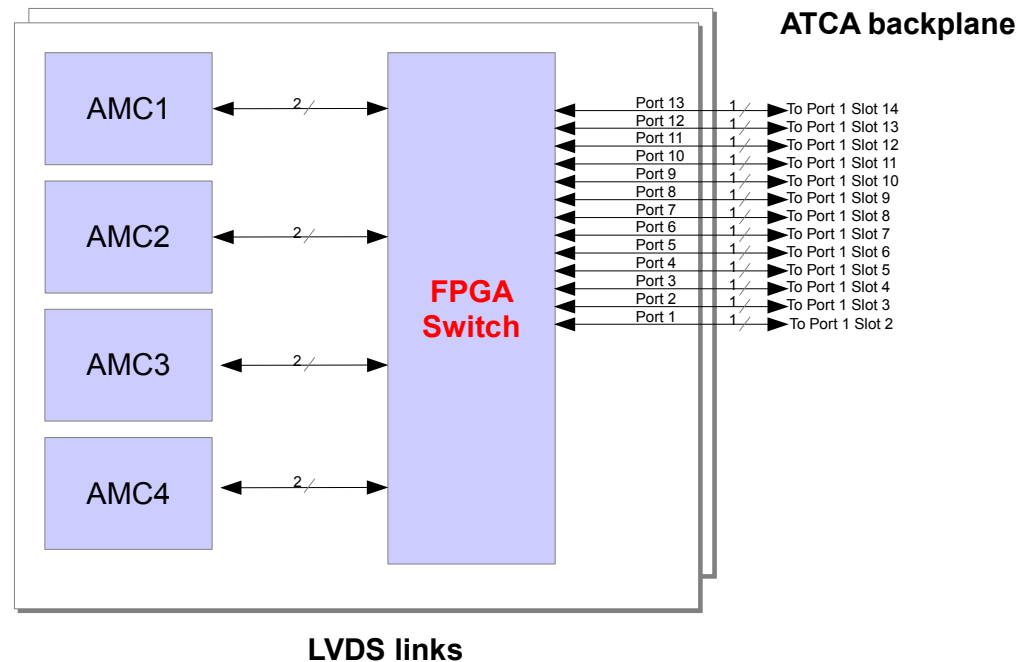
- 72 x 72 CML interface
- Up to 6.5 Gbits/s
- 3 bidir links with each AMC
- 3 bidir links with each of 13 backplane channel
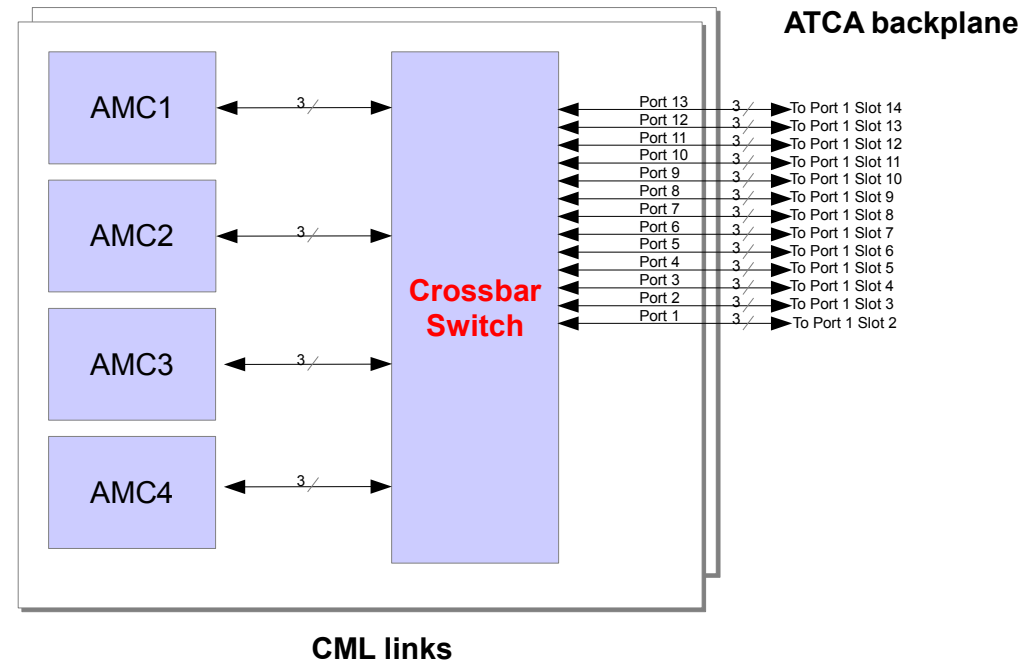
FPGA switch for clocks and throttles
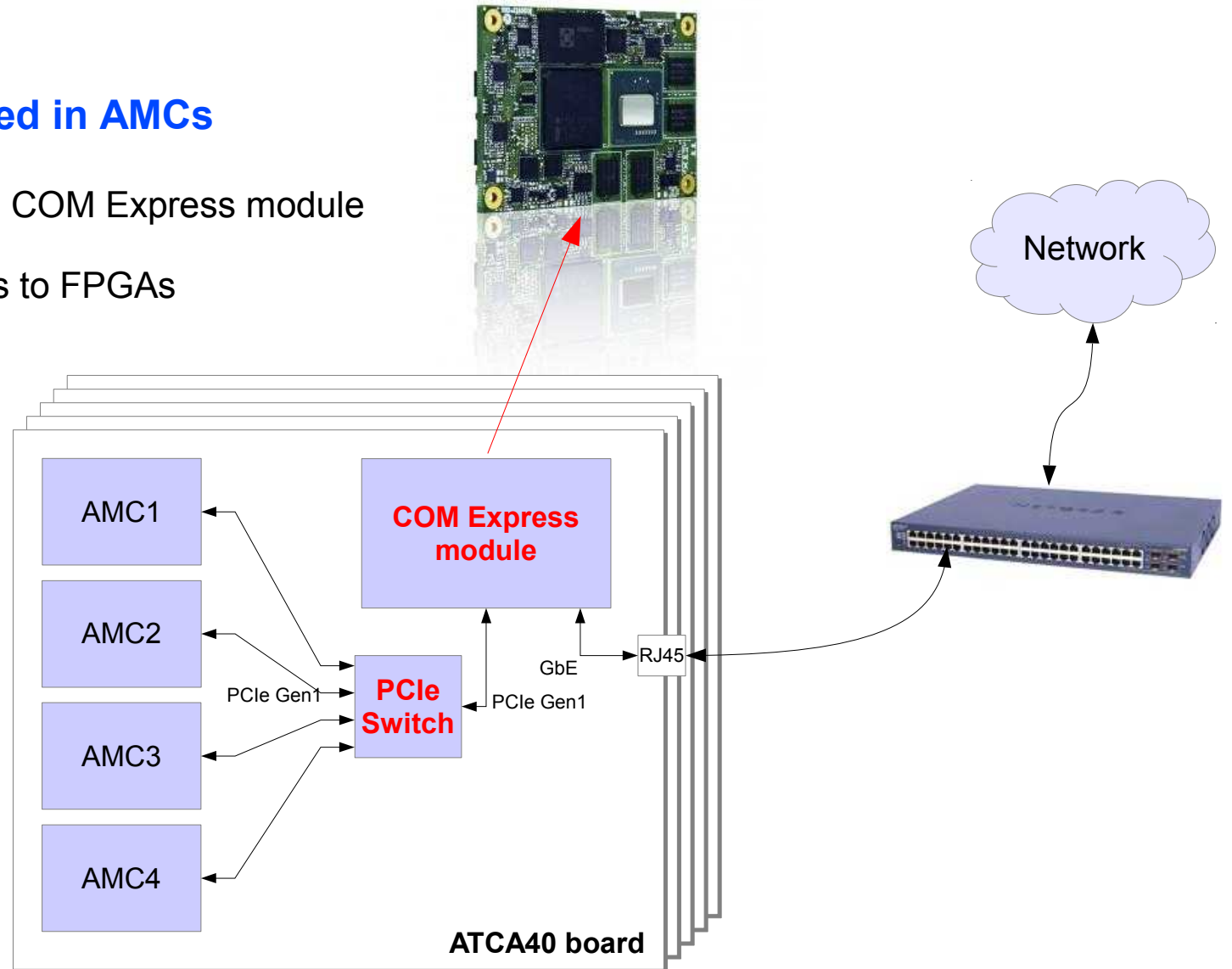
- LVDS interface
- Up to 800 Mbits/s
- 1 bidir clock with each AMC
- 1 bidir throttle with each AMC
- 1 bidir link with each of 13 backplane channel

ATCA backplane

CML links

ATCA backplane

LVDS links

ECS

**From control PC
to FPGAs embedded in AMCs**

- 1 Gbit Ethernet to COM Express module

- 4 PCIe Gen1 links to FPGAs

Acquisition

From Front Ends to Farms

- Output dataflow = Input dataflow

- Direct connection to/from front plate

TELL40

AMC1

24 GBT links at 4.8 Gbits/s

FPGA

12 x 10GbE links at 10 Gbits/s

AMC2

24 GBT links at 4.8 Gbits/s

FPGA

12 x 10GbE links at 10 Gbits/s

AMC3

24 GBT links at 4.8 Gbits/s

FPGA

12 x 10GbE links at 10 Gbits/s

AMC4

24 GBT links at 4.8 Gbits/s

FPGA

12 x 10GbE links at 10 Gbits/s
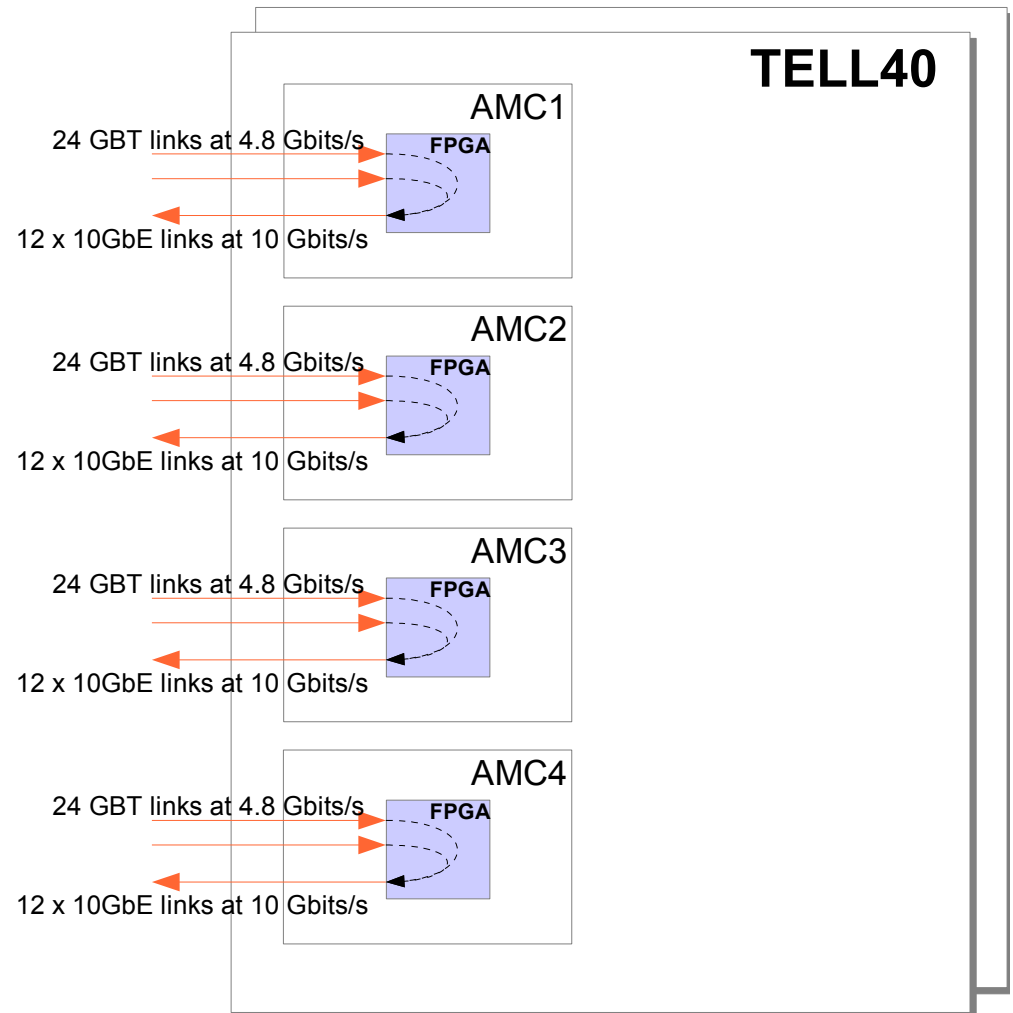
Timing and Fast Control

**From S-ODIN
to FE and TELL40s**

Example of configuration:
many other ones possible …

- Clocks and triggers broadcast over 2.4 Gbits/s serial lines

- Relies on flexibility of Crossbar switch

S-ODIN board

| AMC4 | | FPGA | | AMC1 |
| LHC clock interface | | Switch | | S-ODIN |

AMC_Clock_out
LVDS link, 40 MHz clock
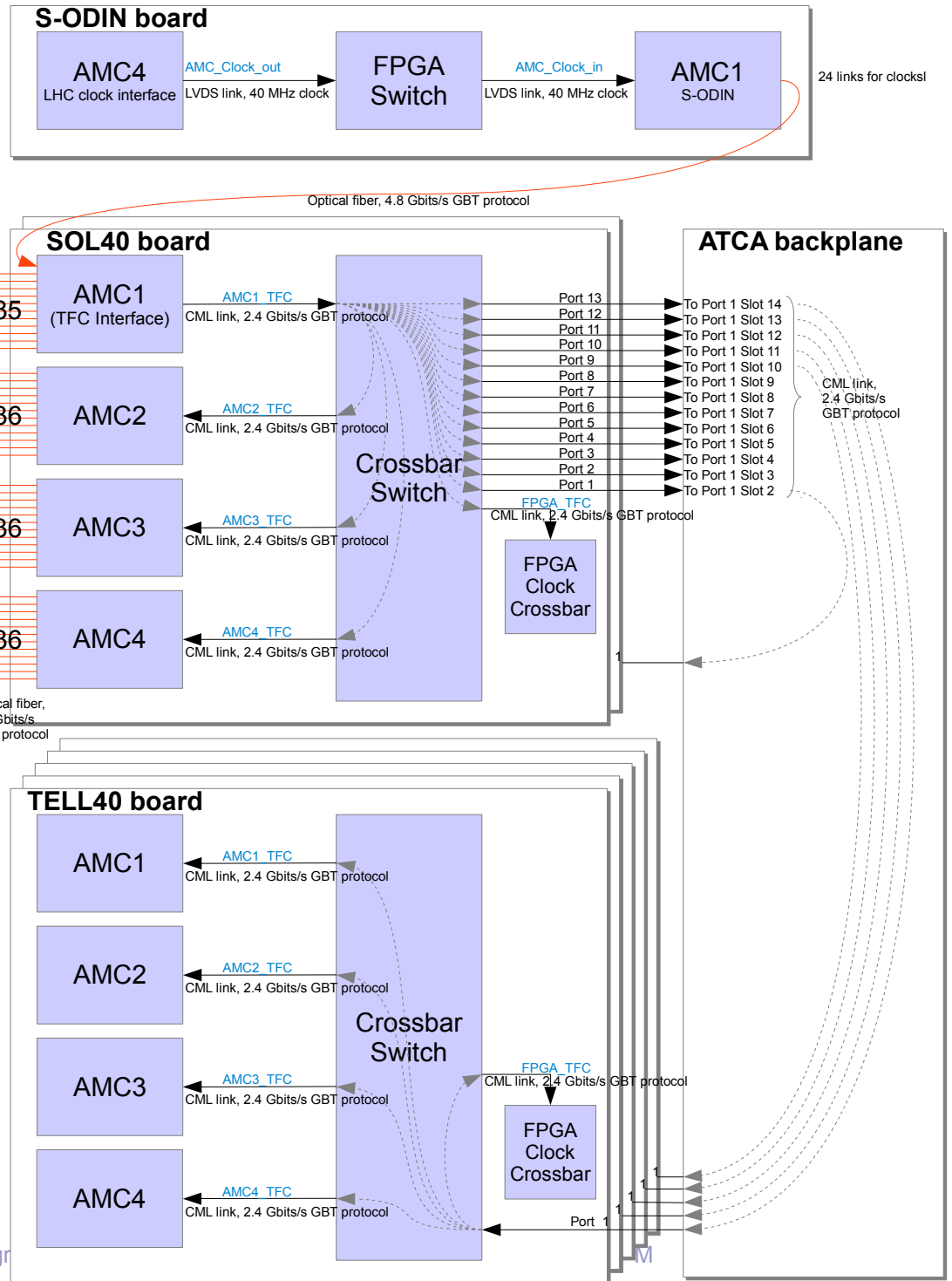
AMC_Clock_in
LVDS link, 40 MHz clock

24 links for clocks!

Optical fiber, 4.8 Gbits/s GBT protocol

SOL40 board

ATCA backplane

AMC1 (TFC Interface) — 35
AMC1_TFC
CML link, 2.4 Gbits/s GBT protocol

AMC2 — 36
AMC2_TFC
CML link, 2.4 Gbits/s GBT protocol

AMC3 — 36
AMC3_TFC
CML link, 2.4 Gbits/s GBT protocol

AMC4 — 36
AMC4_TFC
CML link, 2.4 Gbits/s GBT protocol

Crossbar Switch

Port 13
Port 12
Port 11
Port 10
Port 9
Port 8
Port 7
Port 6
Port 5
Port 4
Port 3
Port 2
Port 1

To Port 1 Slot 14
To Port 1 Slot 13
To Port 1 Slot 12
To Port 1 Slot 11
To Port 1 Slot 10
To Port 1 Slot 9
To Port 1 Slot 8
To Port 1 Slot 7
To Port 1 Slot 6
To Port 1 Slot 5
To Port 1 Slot 4
To Port 1 Slot 3
To Port 1 Slot 2

CML link, 2.4 Gbits/s GBT protocol

FPGA_TFC
CML link, 2.4 Gbits/s GBT protocol

FPGA Clock Crossbar

1

Optical fiber, 4.8 Gbits/s GBT protocol

TELL40 board

AMC1
AMC1_TFC
CML link, 2.4 Gbits/s GBT protocol

AMC2
AMC2_TFC
CML link, 2.4 Gbits/s GBT protocol

AMC3
AMC3_TFC
CML link, 2.4 Gbits/s GBT protocol

AMC4
AMC4_TFC
CML link, 2.4 Gbits/s GBT protocol

Crossbar Switch

FPGA_TFC
CML link, 2.4 Gbits/s GBT protocol

FPGA Clock Crossbar

Port 1
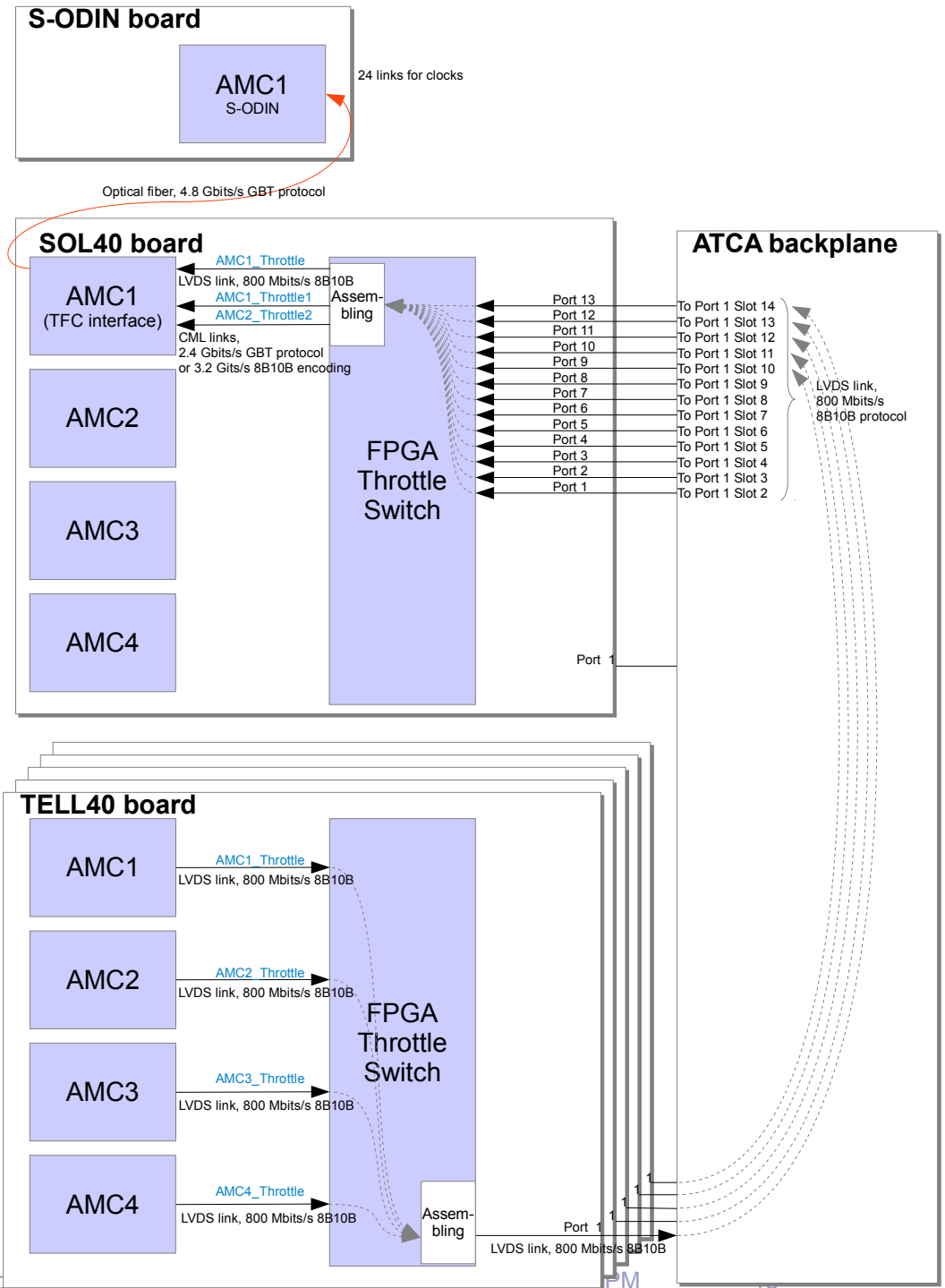
1
1
1

Throttles

From TELL40s to S-ODIN

Example of configuration:
many other ones possible ...

- ◆ Throttles transfered
  over 800 Mbits/s LVDS links
  Information includes
  - B-Id : 12 bits
  - Throttles : 4 bits
  Encoding 8B10B

- ◆ Combination and routing made
 by Cyclone IV FPGAs
  Information includes
  - B-Id : 12 bits
  - Throttles : 52 bits

- ◆ Result transfered to AMC1
  over two 2.4 Gbits/s GBT links

- ◆ Final result sent to S-ODIN
  over one 4.8 Gbits/s GBT link

S-ODIN board

AMC1
S-ODIN

24 links for clocks

Optical fiber, 4.8 Gbits/s GBT protocol

SOL40 board

AMC1
(TFC interface)

AMC1_Throttle
LVDS link, 800 Mbits/s 8B10B
AMC1_Throttle1
AMC2_Throttle2

Assem-
bling

CML links,
2.4 Gbits/s GBT protocol
or 3.2 Gits/s 8B10B encoding

AMC2

AMC3

AMC4

FPGA
Throttle
Switch

Port 13
Port 12
Port 11
Port 10
Port 9
Port 8
Port 7
Port 6
Port 5
Port 4
Port 3
Port 2
Port 1

Port 1

ATCA backplane

To Port 1 Slot 14
To Port 1 Slot 13
To Port 1 Slot 12
To Port 1 Slot 11
To Port 1 Slot 10
To Port 1 Slot 9
To Port 1 Slot 7
To Port 1 Slot 6
To Port 1 Slot 5
To Port 1 Slot 4
To Port 1 Slot 3
To Port 1 Slot 2

LVDS link,
800 Mbits/s
8B10B protocol

TELL40 board

AMC1

AMC1_Throttle
LVDS link, 800 Mbits/s 8B10B

AMC2

AMC2_Throttle
LVDS link, 800 Mbits/s 8B10B

AMC3

AMC3_Throttle
LVDS link, 800 Mbits/s 8B10B

AMC4

AMC4_Throttle
LVDS link, 800 Mbits/s 8B10B

FPGA
Throttle
Switch

Assem-
bling

Port 1

LVDS link, 800 Mbits/s 8B10B

Interface to ODIN

S-ODIN to ODIN:

- Orbit, Clocks, Bid reset … transmitted over LVDS links
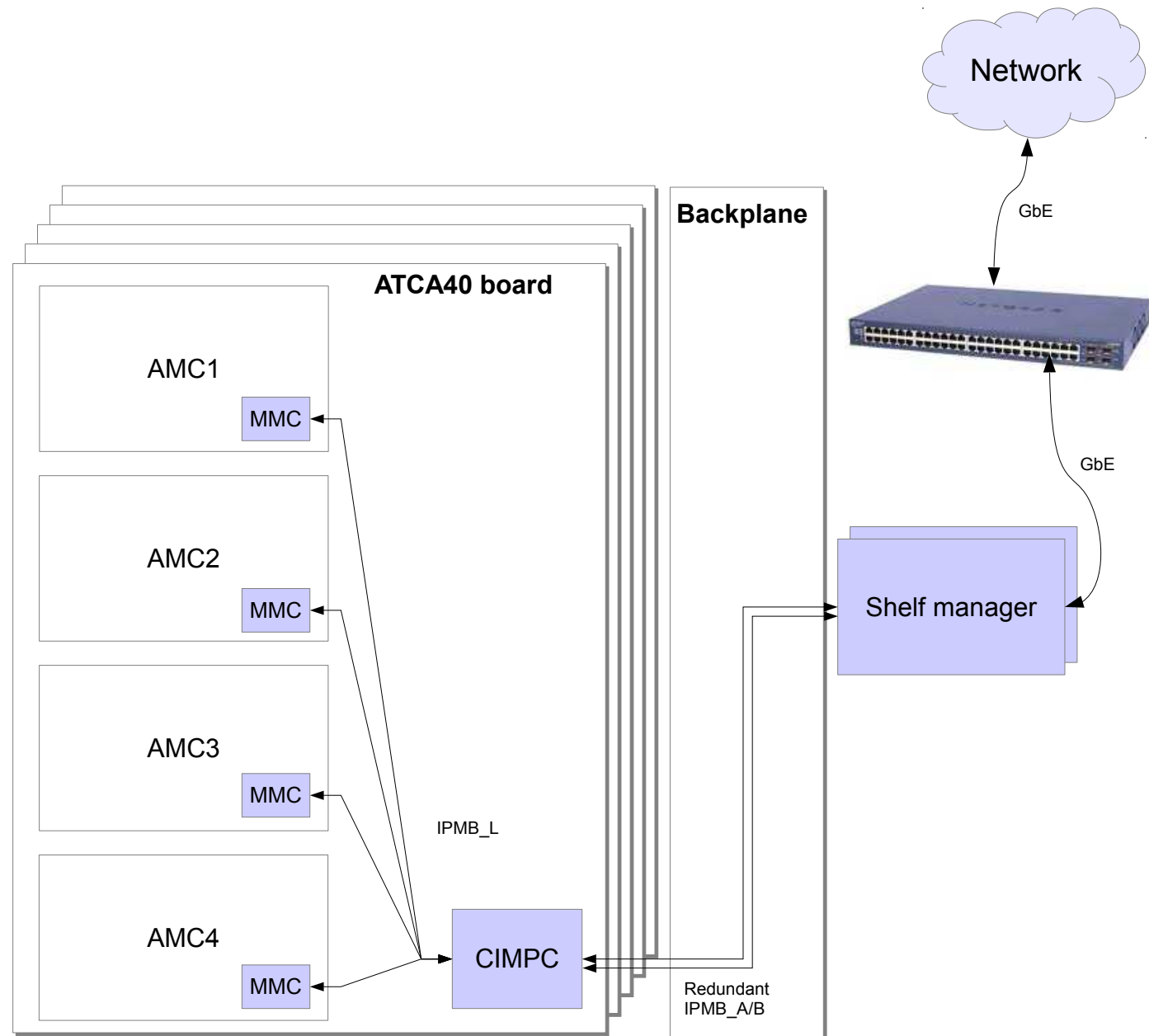
- Conversion on S-ODIN RTM board in ECL signals
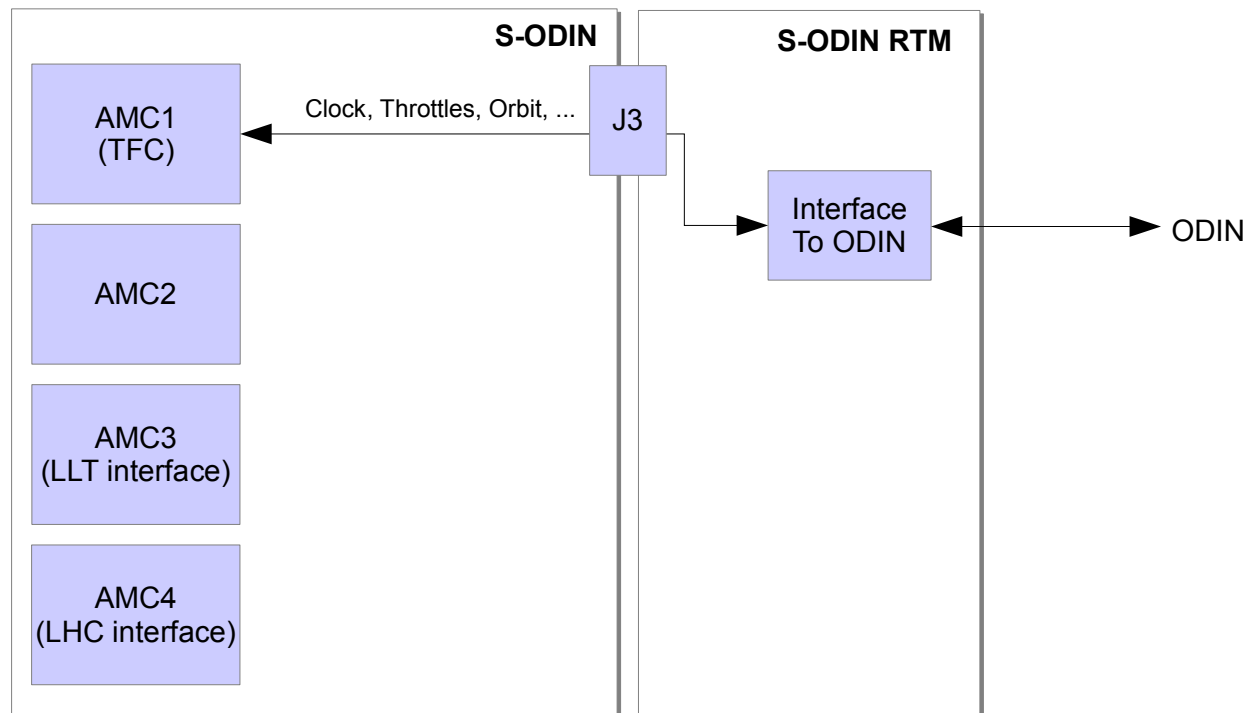
IPMI data path

From control PC to MMCs

- 1 Gbit Ethernet to Shelf Manager

- Redundant IPMB A/B to CIPMC

- IPMB_L to MMCs

Functions limited to:

- Hard resets

- Temperature, voltage monitoring

- Switching on and off the board

Network

Backplane

ATCA40 board

AMC1

MMC

AMC2

MMC

AMC3

MMC

IPMB_L

AMC4

MMC

CIMPC

GbE

GbE

Shelf manager

Redundant IPMB_A/B

Backplane

Crate, power consumption, shelf manager ...

Use of a 14 slots Schoff crate

- Lab test purpose for the time being
- Mechanical adaptation needed
  to reuse vertical heat exchangers
- Full-Mesh topology

Power consumtion

- Estimated power consumption per board :
  ~ 250 W → to 3.5 kW /crate
- Crate cooling capacity: 4.2 kW

Shelf manager

- Pigeon Point ShMM-ACB-V
- Chosen for compatibility reasons
  with LAPP IPMI board

Schroff 14 slots crate
(ref 11596-30x)

Feasibility

Clock phase over serial links (1)

Can we keep a fix clock phase over chained serial links when powering on and off the system ?

LHCb Upgrade
FPGA — TFC — A (First hop) — FPGA — TFC Interface (TELL 40) — B (second hop) — GBT Front End
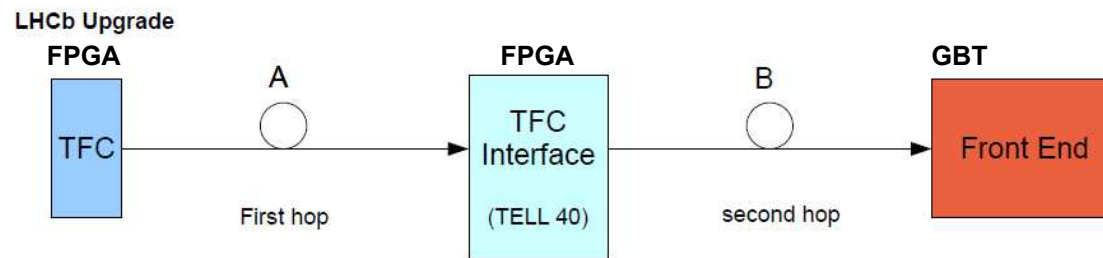
- Recovered clock phase is not constant in a FPGA deserializer

Succesfully tested on Stratix IV GX

- Use of a deterministic latency mode
- Phase variation on second serial stream :
  ~ ± 50 ps RMS
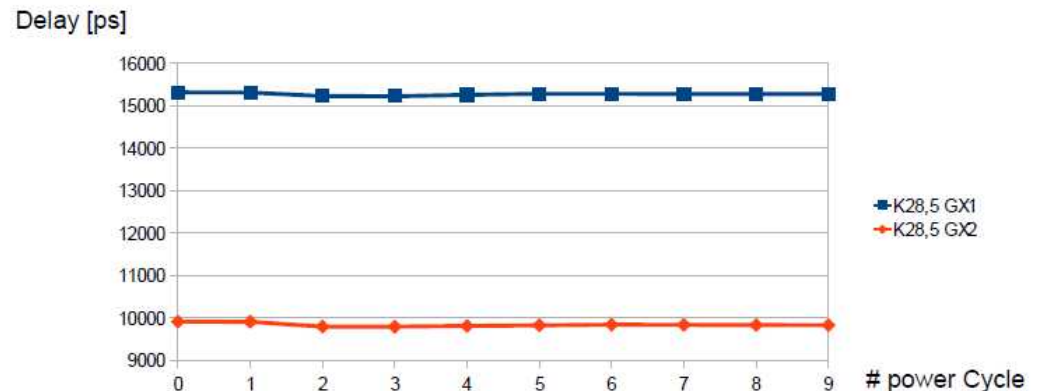
Clock phase over serial links (2)

Preliminary results : need more statistics

- Next steps

  - Port this design on Stratix V GX

  - Measurements from chip to chip through backplane with several hops

  - Implement equivalent mechanism on GBT protocol
    - *Experimental on-going work by Federico Alessio and Richard Jacobsson*
      Use of a not interleaved GBT protocol at 2.4 Gbits/s

Optical links

BER < 10^{-16} at 4.8 Gbits/s and 10.3125 Gbits/s
over 10 meters OM3 optical fiber
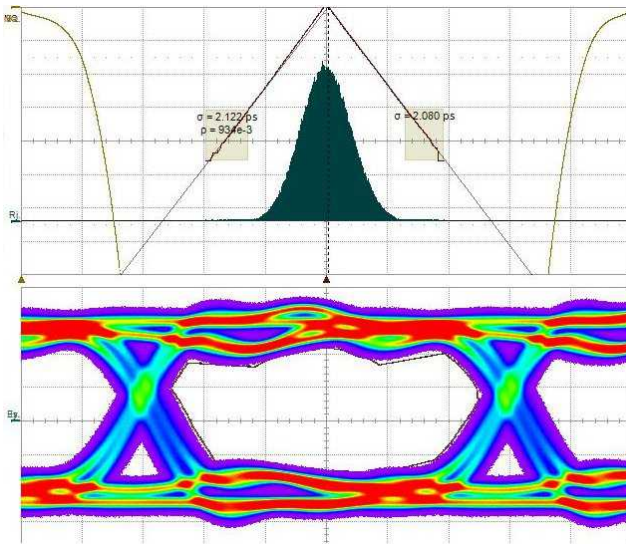
Measurements at 4,8 Gbit/s :
Total Jitter   ≈ 56 pS
Random Jitter   ≈ 2,4 pS
Deterministic Jitter ≈ 24 pS
aperture : 0,65 UI@10^{-16}
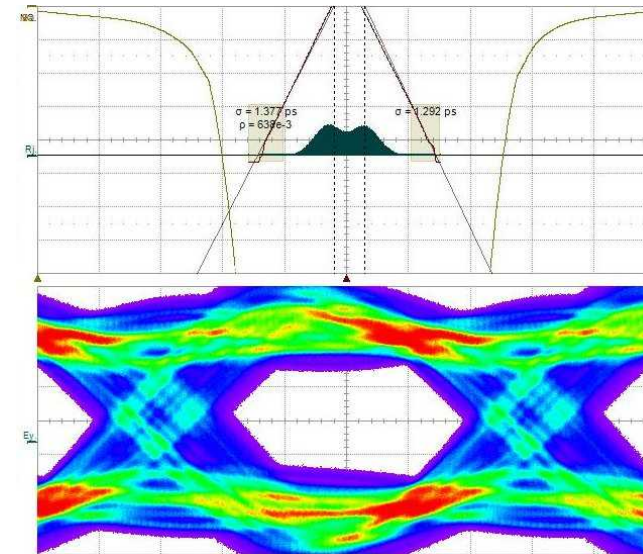
Measurements at 10.3125 Gbit/s :
Total Jitter   ≈ 55 pS
Random Jitter   ≈ 0.93 pS
Deterministic Jitter ≈ 42 pS
aperture : 0,42 UI@10^{-16}

Further qualification required:
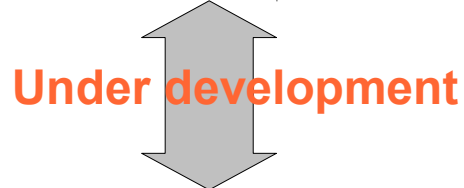- *exhaustive test of all links of the board,*
- *crosstalk,*
- *100, 300, 400 meters fibers*

IPMI risk mitigation

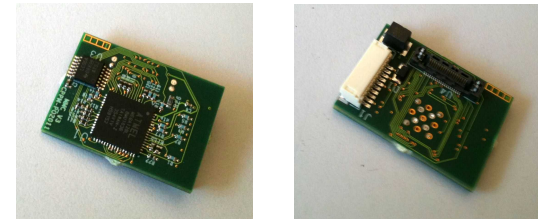**Use of an open source and already tested chain**

- Shelf manager Pigeon Point

Tested

- CIPMC developped by Annecy
  - ➜ *Based on open source CoreIPM*

Under development

- MMC developped by DESY/CPPM/CERN
  - ➜ *Basé sur open source DESY*

CIPMC

Mezzanine MMC

Common solution with ATLAS

Firmware development

# Motivations for a Low Level Interface

Many firmwares but common data flows and requirements

- Same data path for control
- Same input and output data path for processing
- Need for embedded stand-alone simulators (FE, TFC, LLT, Farm, ...)
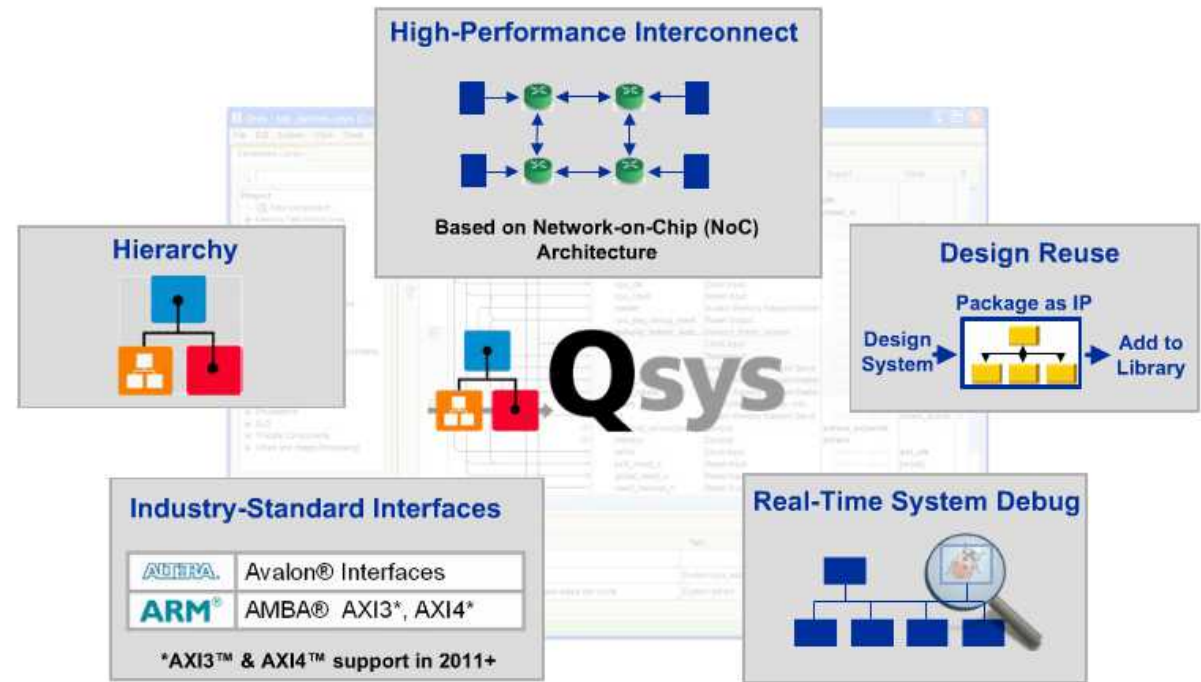- Monitoring buffers

Need for a flexible low level interface in which user code can be "plugged"

- Hide the underlying complexity (GX buffers, GBT, PCIe, 10 GbE, …)

Marseille proposal : QSYS
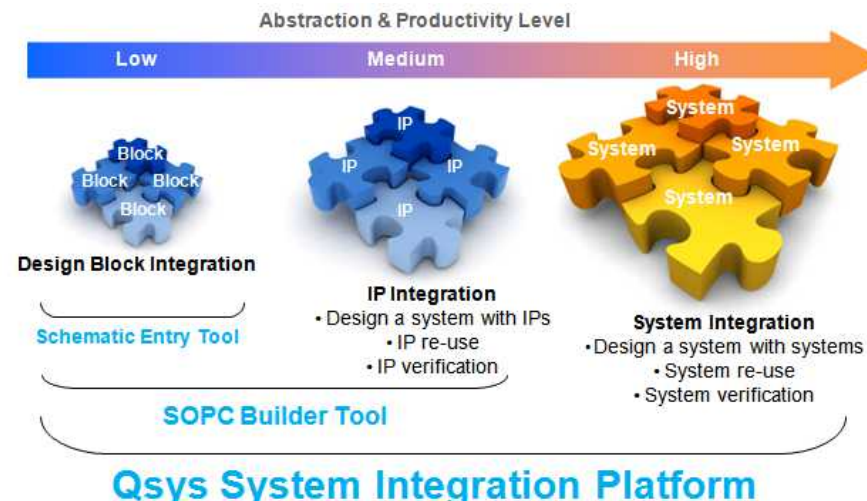
QSYS features

Powerful system integration tool

- High level of abstraction for design capture
- Facilitates design reuse

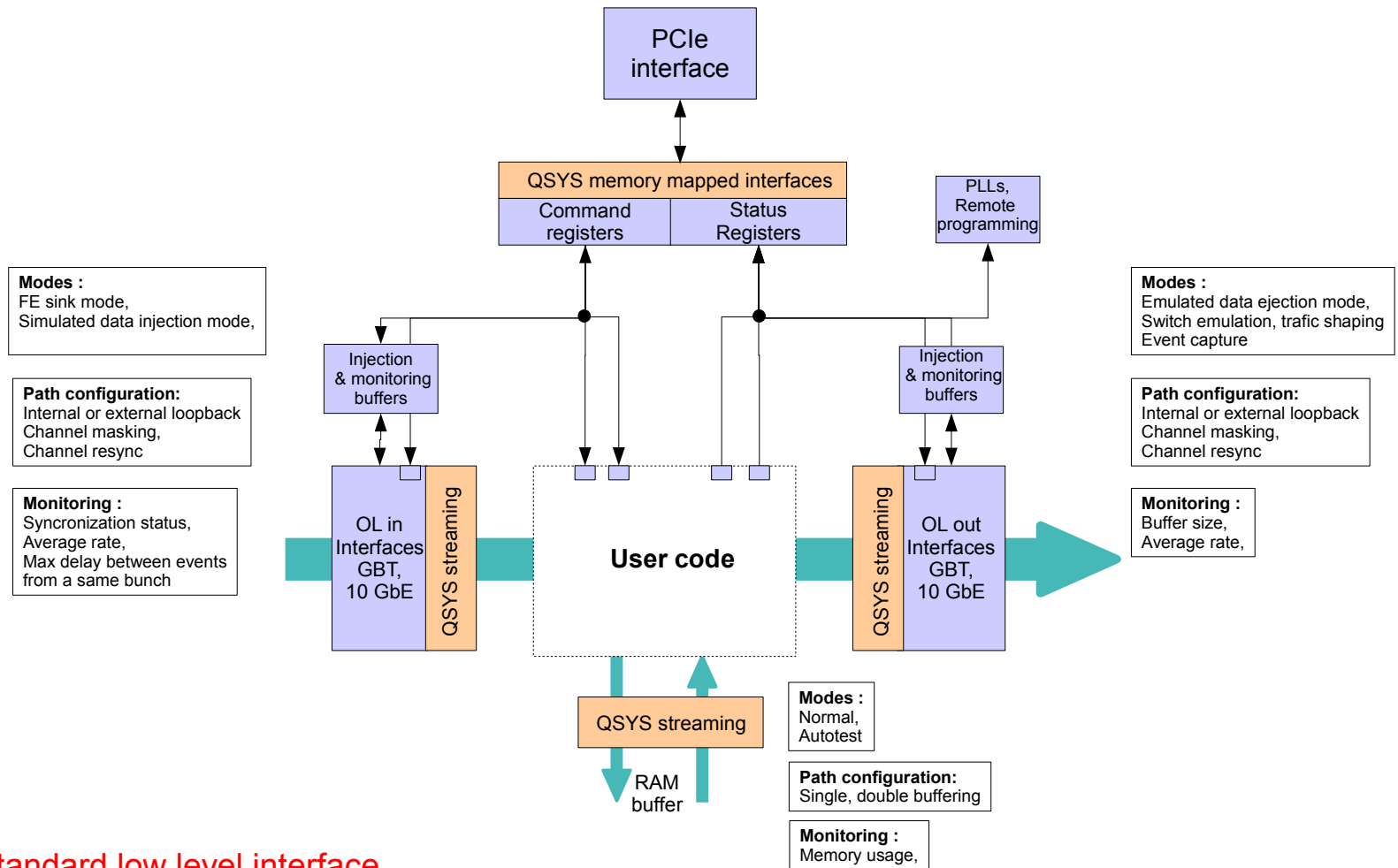## Save time by avoiding writing HDL code for interconnection

- Automatically creates high-performance interconnect logic

Easy way to normalize interfaces in the system

- Standard interfaces
- Documentation maintained by Altera and already available

Automatic test bench generation

Low level Interface

PCIe interface

QSYS memory mapped interfaces

Command registers | Status Registers

PLLs, Remote programming

Modes :
FE sink mode,
Simulated data injection mode,

Path configuration:
Internal or external loopback
Channel masking,
Channel resync

Monitoring :
Syncronization status,
Average rate,
Max delay between events
from a same bunch

Injection & monitoring buffers

Injection & monitoring buffers

OL in Interfaces GBT, 10 GbE

QSYS streaming

User code

QSYS streaming

OL out Interfaces GBT, 10 GbE

Modes :
Emulated data ejection mode,
Switch emulation, trafic shaping
Event capture

Path configuration:
Internal or external loopback
Channel masking,
Channel resync

Monitoring :
Buffer size,
Average rate,

QSYS streaming

RAM buffer

Modes :
Normal,
Autotest

Path configuration:
Single, double buffering

Monitoring :
Memory usage,

- Standard low level interface will be provided to firmware developpers

Conclusion

Flexible architecture

- Single hardware, easy to maintain
- Modular and reconfigurable
- Can take into account future needs

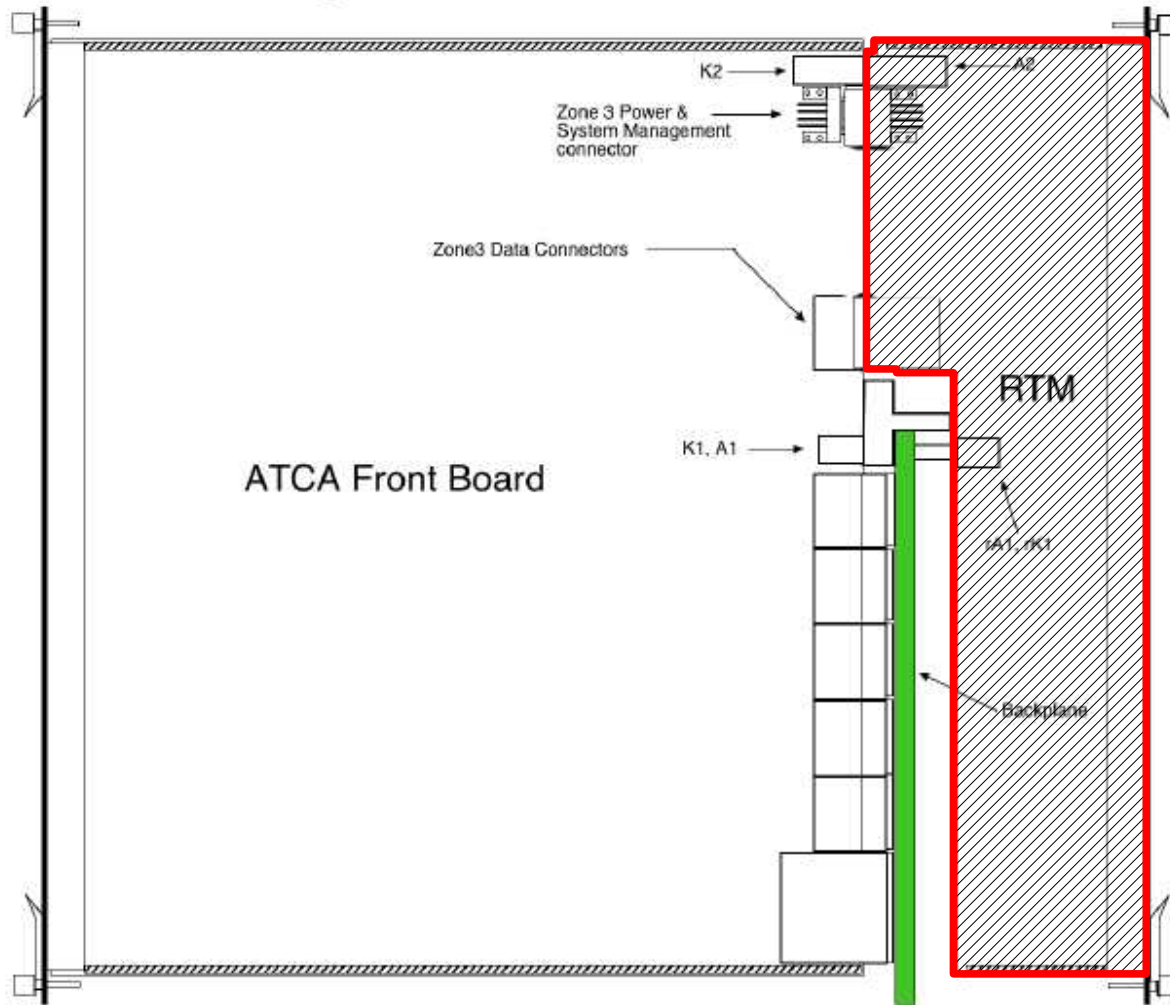**Early measurements make us confident on feasibility**

- Clock phase stability over serial link: ~ ± 50 ps RMS
- High speed BER < 10^{-16} at 10.3125 Gbits/s

Distributed development with common LLI

Ongoing validation of a full scale prototype

Backup slides

S-ODIN RTM

K2

Zone 3 Power &
System Management
connector

Zone3 Data Connectors

RTM

ATCA Front Board

K1, A1

A1, K1

Backplane

RTM Interface to be defined
Connexions already present on ATCA40

FPGA choice

Package	Stratix V GX								Stratix V GT		Stratix V GS					Stratix V E	
	A3	A4	A5	A7	A9	AB	B5	B6	C5	C7	D3	D4	D5	D6	D8	E9	EB
EH29-H780	✓										✓	✓					
HF35-F1152 [2]	✓	✓	✓	✓							✓	✓	✓				
KF35-F1152	✓	✓	✓	✓													
KF40-F1517 / KH40-H1517	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓		
NF40 / KF40-F1517 [3]			✓	✓					✓	✓							
RF40-F1517							✓	✓									
H40-H1517																✓	✓
RF43-F1760							✓	✓									
NF45-F1932			✓	✓	✓	✓								✓	✓		
F45-F1932																✓	✓

Notes to Table 1–5:

(1) All devices in a given row allow migration.

(2) All devices in this row are in the HF35 package and have twenty-four 14.1-Gbps transceivers.

(3) The 5SGTC5/7 devices in the KF40 package have four 28.05-Gbps transceivers and thirty-two 12.5-Gbps transceivers. Other devices in this row are in the NF40 package and have forty-eight 14.1-Gbps transceivers.

Links repartition on an AMC board

	TFC+ECS	TELL40 GBT Frame Format	TELL40 GBT Wide bus mode
Input protocol	GBT	GBT	GBT
Output protocol	GBT	10 GbE	10 GbE
Input links	up to 36	24	24
Output links	up to 36	8 (up to 12)	12

Links organization   Links organization

Data paths : Acquisition
Timing and Fast Trigger
Slow control

Clock phase over serial links

Can we keep a fix clock phase over chained serial links when powering on and off the system ?
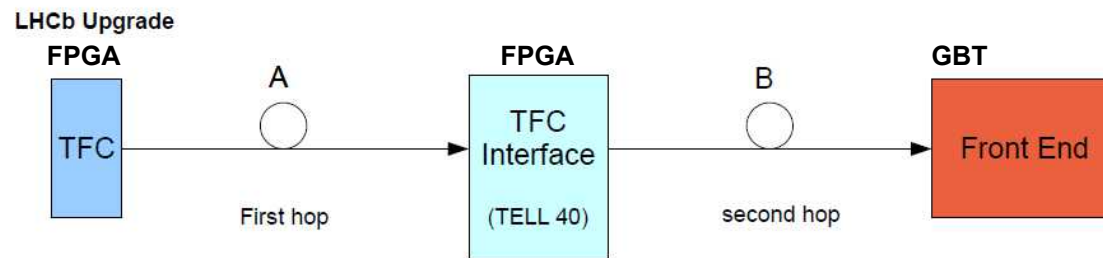
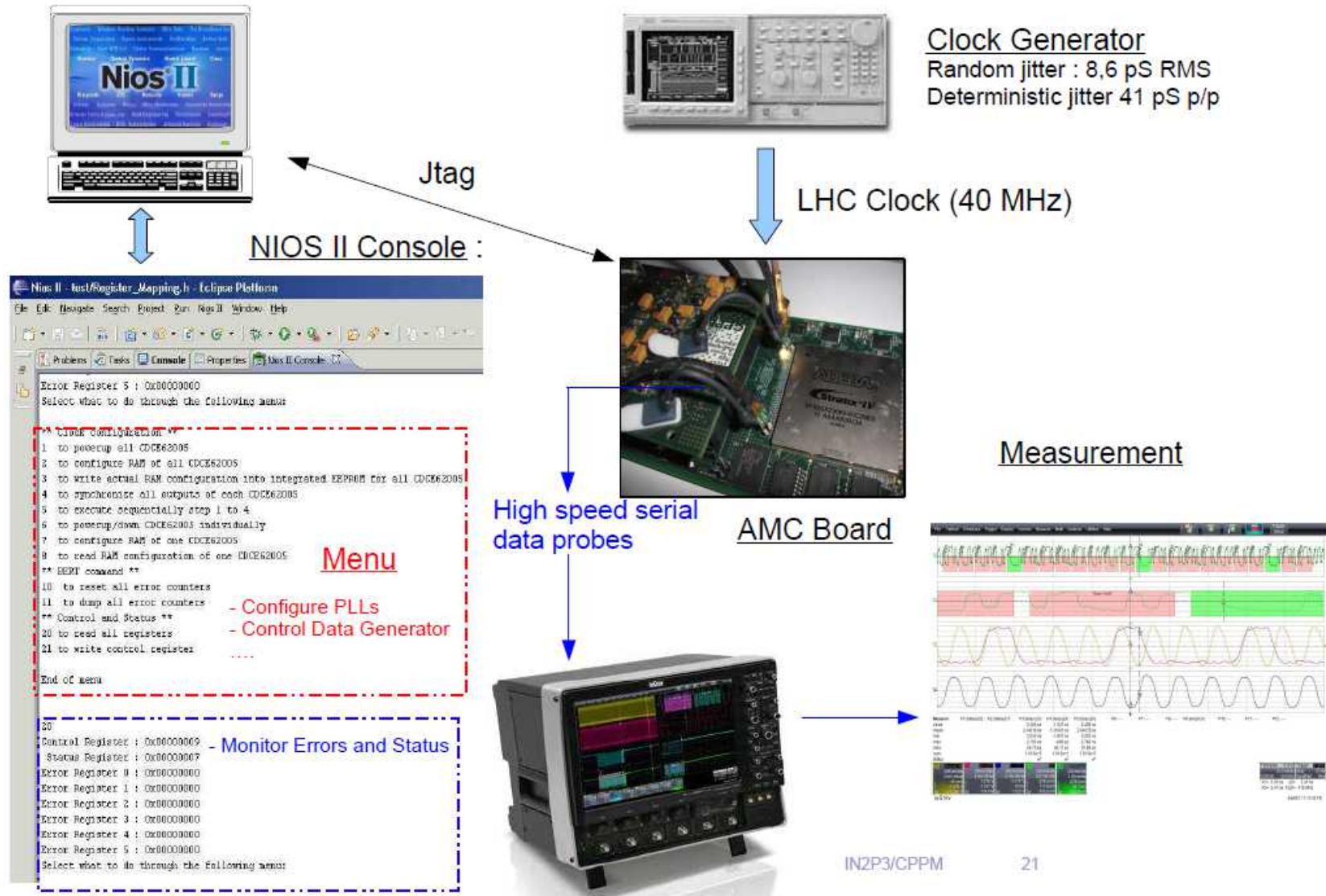- Recovered clock phase is not constant in a FPGA deserializer
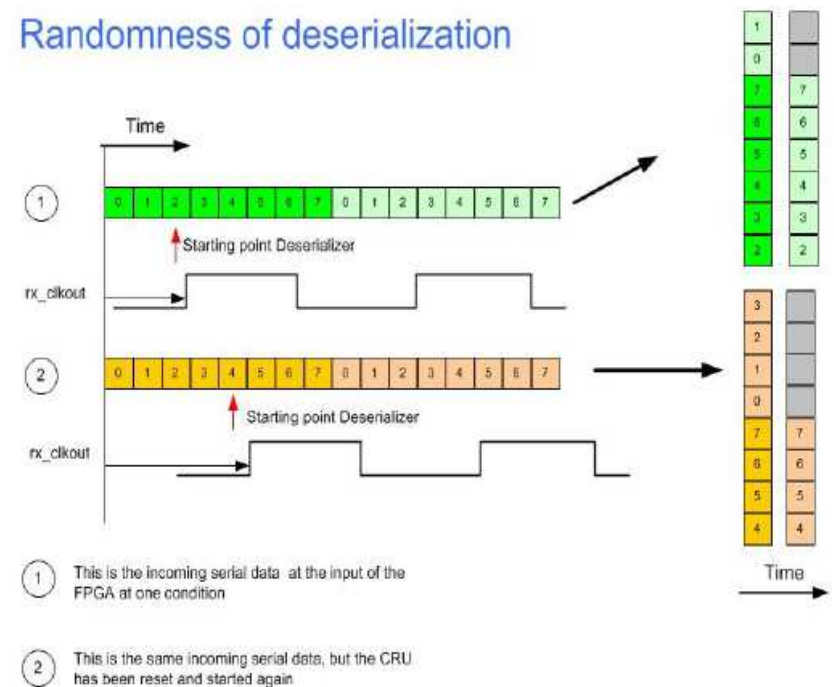
Methodology

Special mechanisms present in Stratix IV GX and V GX can be used
- Early tests made last year with Stratix IV GX
- Use of 8B10B code rather than GBT protocol to be able to detect the phase over a serial link
- Serial link speed : 2.4 Gbits/s
- Whole path emulated in a **single FPGA**

Test Set Up on AMC board

Data Generator → GXB1 TX → A → GXB1 RX → FIFO → GXB2 TX → B → GXB2 RX → Data Test

Recovered clock

first GX buffer stage | second GX buffer stage | **Single FPGA**

CDCE62005
External jitter Cleaning PLL

TFC emulation | **TFC interface emulation** | **Tell40 or Front-End emulation**
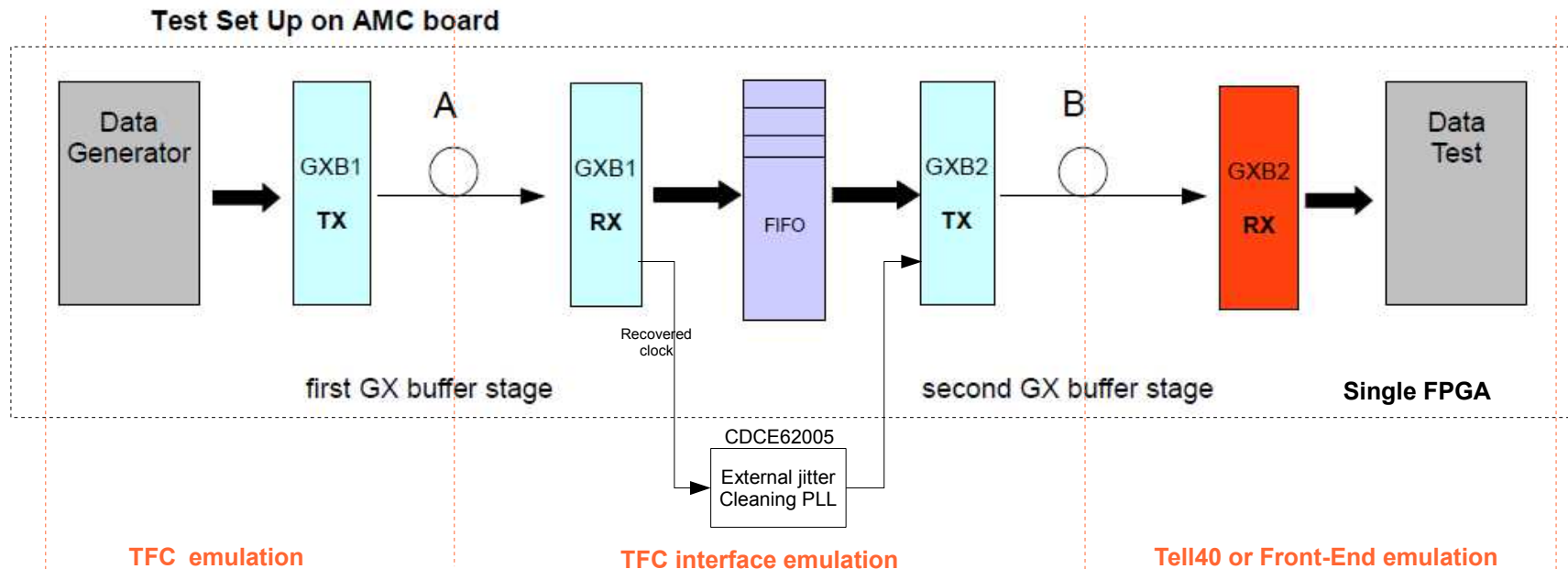
Test setup

Measurements : Set UP

Clock Generator
Random jitter : 8,6 pS RMS
Deterministic jitter 41 pS p/p

Jtag

LHC Clock (40 MHz)

NIOS II Console :

Measurement

High speed serial
data probes

AMC Board

Menu
- Configure PLLs
- Control Data Generator
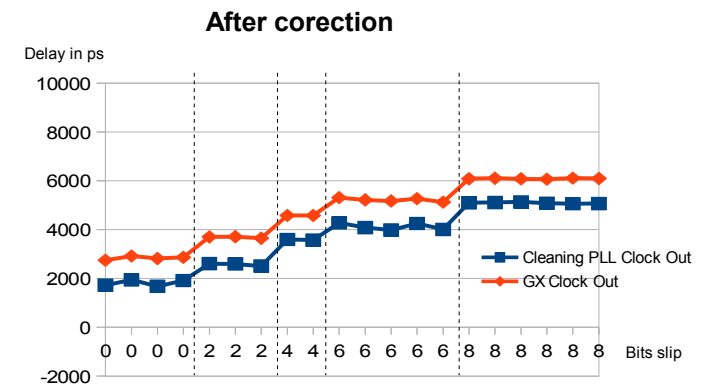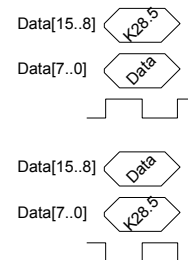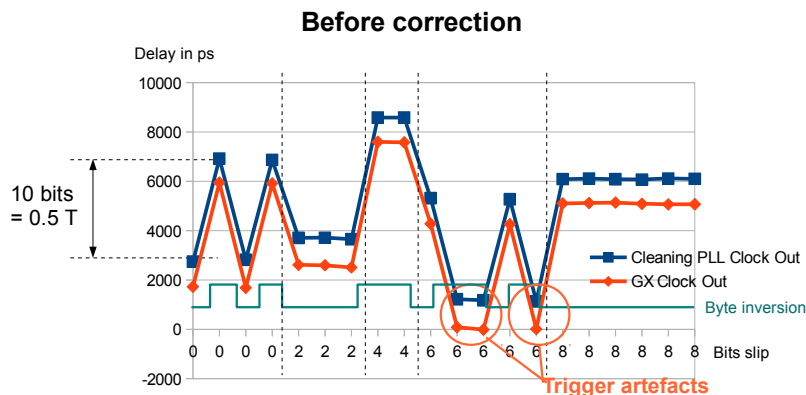....

- Monitor Errors and Status

IN2P3/CPPM   21

Results

Phase of serial stream vs Tx_Clock

- Deterministic delay between clock and serial stream when powering on and off
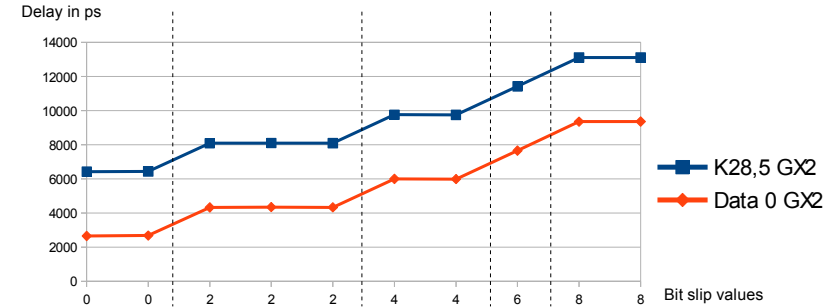- Phase variation: ± 50 ps

Delay [ps]

Delay from trigger to « Sync pattern » (K28,5)
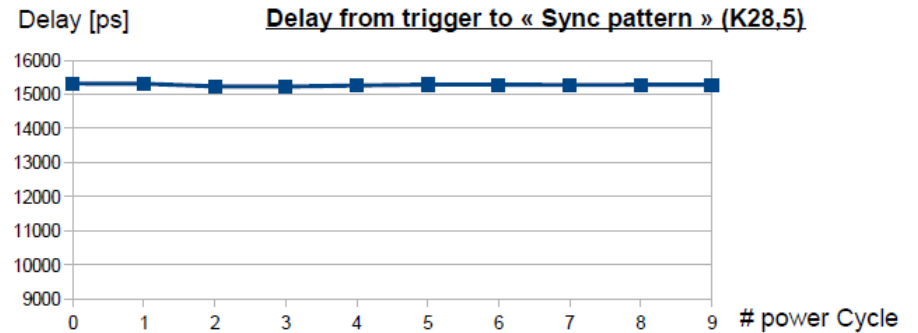
Phase of recovered clock vs Tx_Clock

- Delay = deterministic function of Bit slip out information from receiving GX + Control flags

Before correction

Delay in ps

10 bits = 0.5 T

Cleaning PLL Clock Out
GX Clock Out

Byte inversion

Bits slip

Trigger artefacts

Data[15..8] K28,5
Data[7..0] Data

Data[15..8] Data
Data[7..0] K28,5

After corection

Delay in ps

Cleaning PLL Clock Out
GX Clock Out

Bits slip
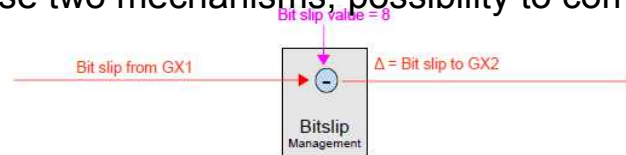
Locking the phase of two serial steams

Phase of serial stream vs bit_slip_in

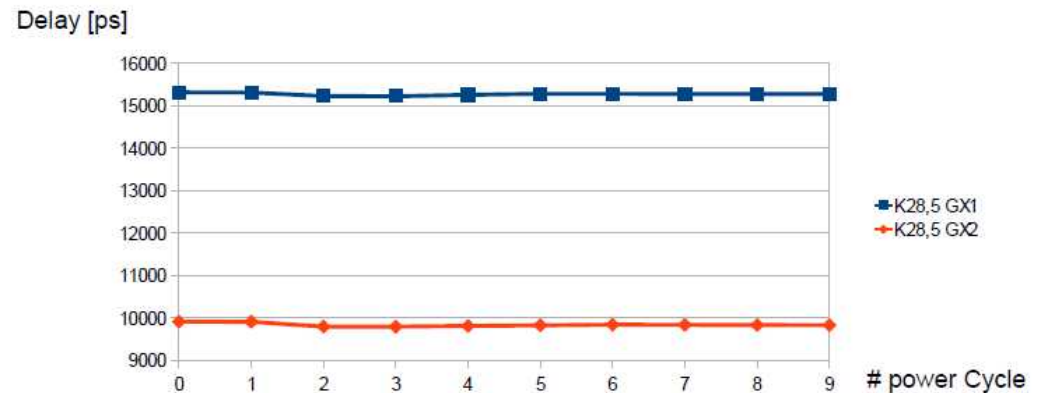- Deterministic relationship

Phase of serial stream 2 vs serial stream 1

- With these two mechanisms, possibility to compensate before resending data

- Phase variation on second serial stream after compensation :
  ~ ± 50 ps RMS

Hardware resets