Next generation readout platform for future detectors :
R&D PCIe400 team

Julien Langouët *(CPPM) on behalf of the R&D PCIe400 team*
CPPM, IJClab, IP2I, LAPP, LPCC, LHCb Online

Outline

- Current development : PCIe400
- Plans for future
- Foreseen evolution based on experience
- Synthesis

Current development : PCIe400

R&D PCIe400 project context

Goals

- Develop a generic readout DAQ card interfacing up to 48 links for custom protocol (GBT/lpGBT) to 1 commercial protocol high bandwidth link (PCIe Gen5/400GbE)
- Cope with tighter timing requirement of upcoming detectors for time fast control (TFC)
- Explore experimental path such as integrating a network interface and complex data processing

Target deployment of PCIe400 is during LS3 for upgraded detectors

- Interest from LHCb, Belle II, CTA and Alice collaborations

IN2P3 R&D

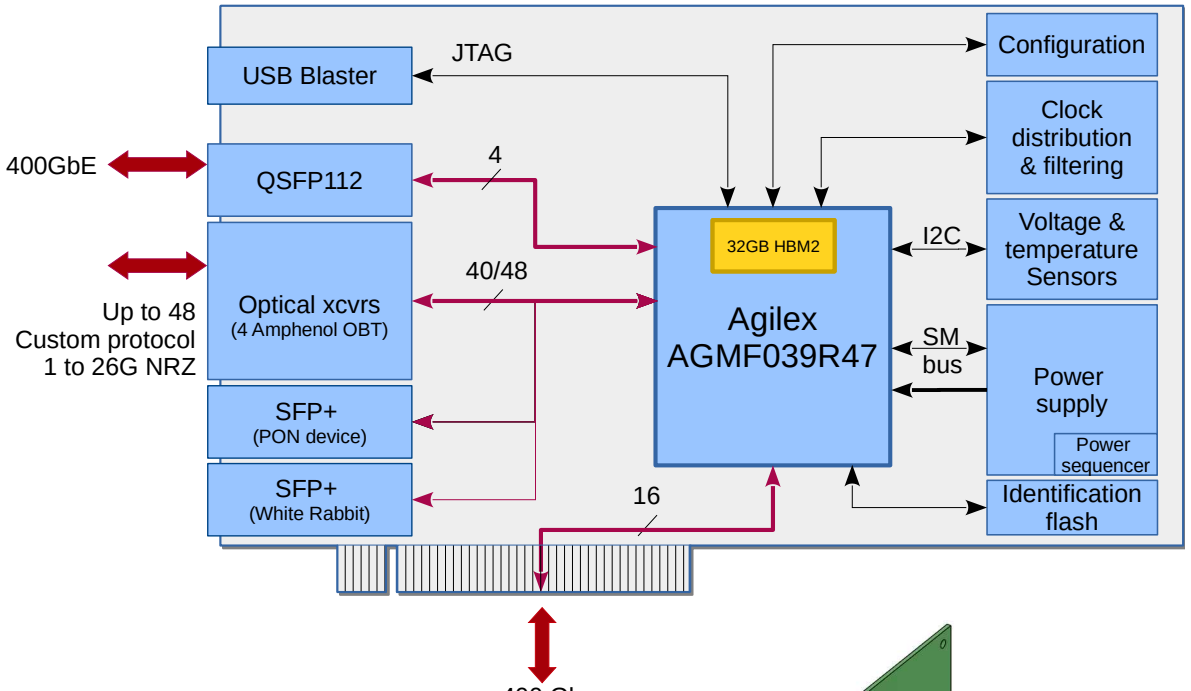
- Project set up to develop the prototype of PCIe40 next generation
- Funded for 3 years from 2022 to end of 2024
- Unite the workforce of 5 labs from IN2P3 as well as LHCb online team @CERN

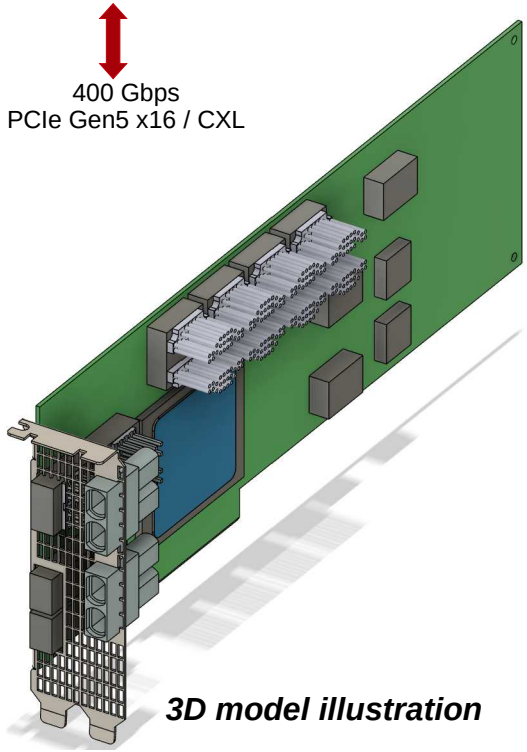
PCIe400

Foreseen characteristics

- PCIe Add in Card 3/4 length
- Agilex 7 M-series AGMF039R47A1E2V
  - Processing capabilities x8 - 12 compared to previous generation FPGA (Arria 10)
- No DDR memory
  - Use of server RAM or HBM2e instead
- Up to 48x26Gbps NRZ for FE
- PCIe Gen 5 / CXL
- QSFP112 for 400GbE (experimental)
- 2 SFP+ for White Rabbit clock distribution or PON fast control
- High precision PLLs jitter <100fs RMS with phase control

PCIe400 synoptic



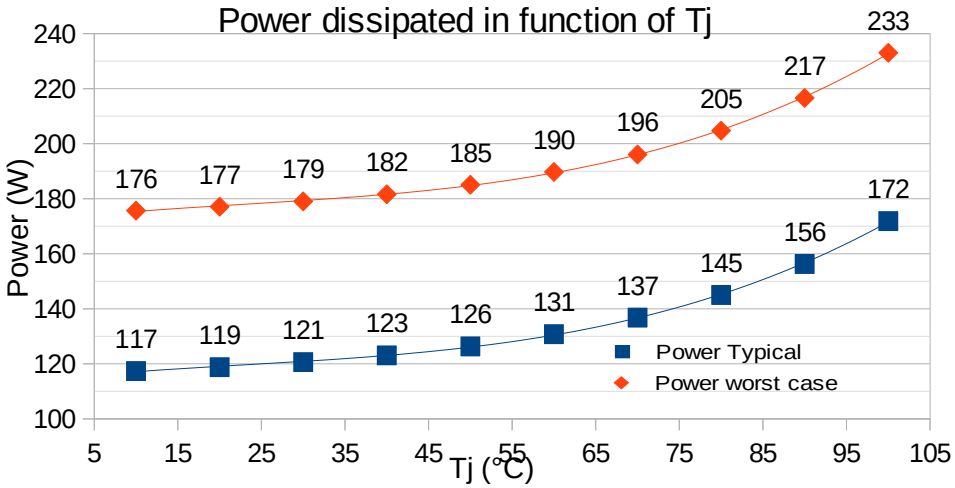


3D model illustration

Power dissipation

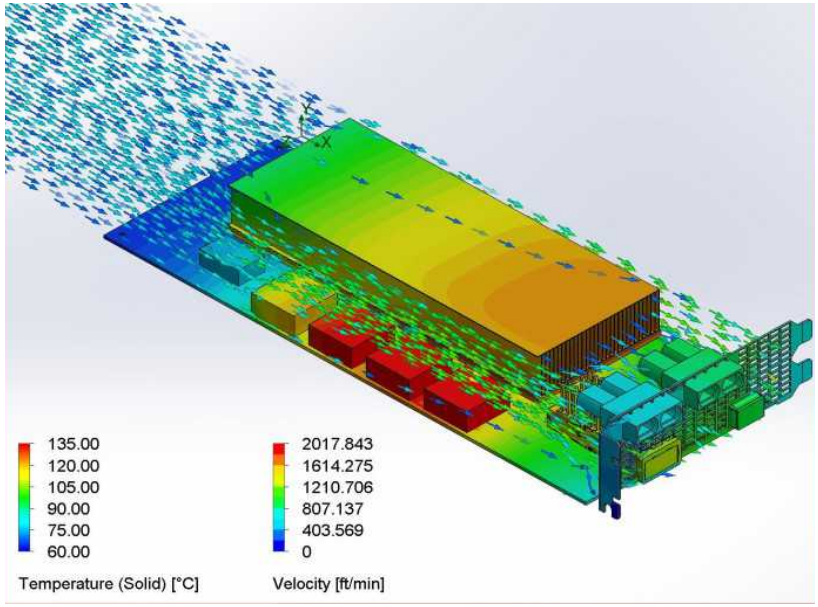
FPGA total power dissipated (TDP)

- Estimated between 120W to 230W
- Up to 100A current for FPGA core
- Need for high performance cooling solution



Cooling solution

- CFD simulations to study air cooling feasibility with vapor chamber heatsink
- Optical transceiver are the determining factor because of cumulative heat along airflow effect



Optical interface

4x Amphenol OBT

- 12 duplex channels (MPO-24)
- 1.25G to 26.3G NRZ
- Specs are compatible with VTRx+

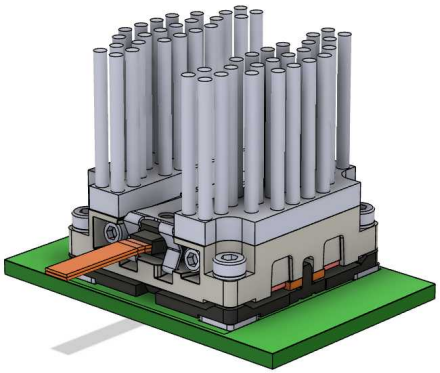
2x SFP+ 10G for TFC / White Rabbit

QSFP112

- 400GbE (4x112G PAM4)
- Direct Attach Cables are available <3m
- Optical modules slowly become available

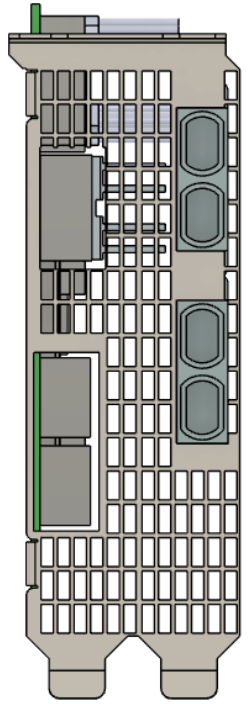
Limited number of FPGA transceiver

- Number of serials links depend on configuration



Amphenol OBT
1.25G à 26.3G NRZ

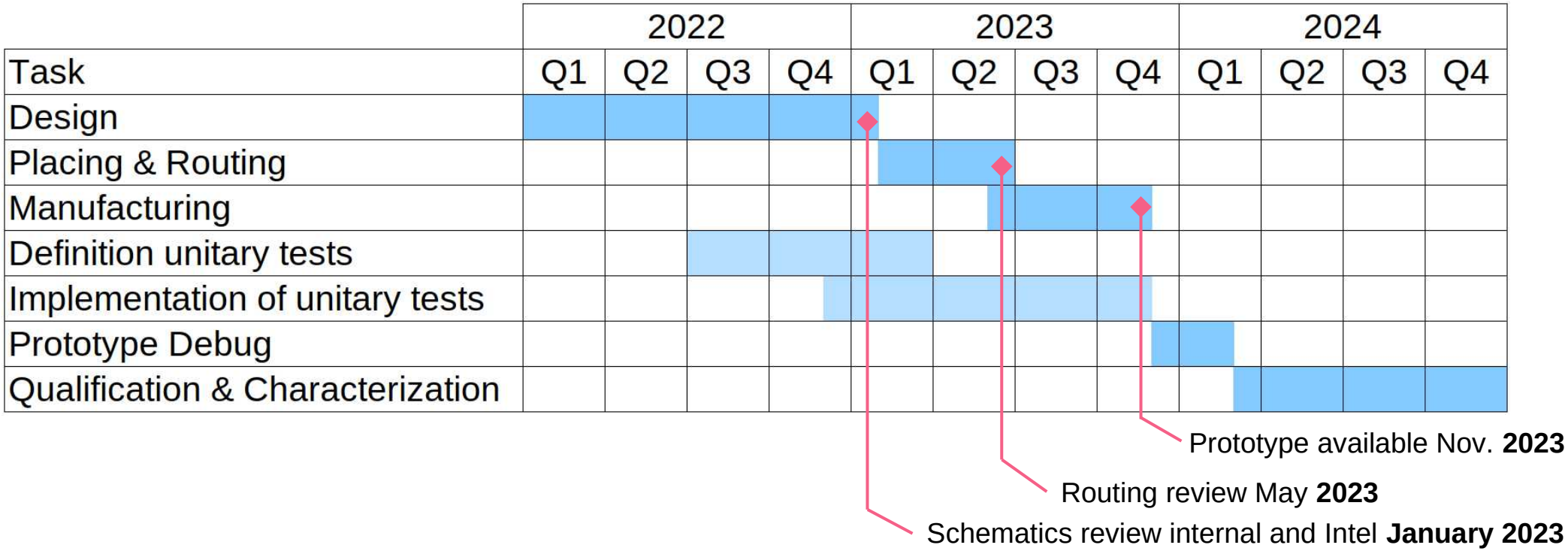
QSFP112
106.25Gb/s PAM4



PCIe400 front-view

	# FE links
No TFC/WR/400GbE	48
WR	47
TFC (TTC-PON)	46
TFC (TTC-PON) + 400GbE	38

Planning



Placement and routing specification finished

- Draft PCB stackup to refine with manufacturer
- Hardware simulation for power and signal integrity planned during routing phase

Software and firmware developped and test in parallel with devkits

Plans for future

Exploring Network interface on-board

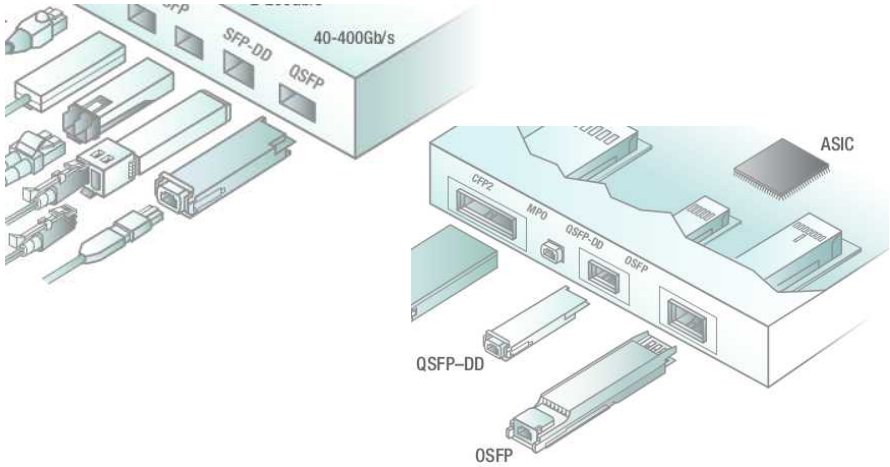
Current FPGA capabilities

- Dedicated ‘FHT’ interface for high datarate and 400GbE hard IP with MAC, PCS and FEC layers
- 32GB HBM2e memory and Network on-chip hard IP (NoC) for data moving and buffering

Datapath Clocking Mode	Configuration	Data Rate R
		FHT PMA
PMA clocking mode (maximum 906.25 MHz)	PMA Direct	• 24-29 Gbps NRZ • 48-58 Gbps NRZ and PAM4 • 96-116 Gbps PAM4
System PLL clocking mode (maximum 1 GHz)	PMA Direct Other configurations with MAC, PCS, and FEC	

QSFP112 form factor selected for PCIe400

- QSFP112 is natural evolution from 200G QSFP56 and backward compatible with QSFP
- Best compromise to allow both a symetrical 48 links through MPOs and additional 400GbE interfaces



Technical challenges

- Route differential pairs 112Gb/s PAM4 over $\mathcal{O}$(10)cm
- Implement RDMA over Converged Ethernet (RoCE) network stack in FPGA

Application

- Get rid of NIC in data acquisition path
- Build network of PCIe400 for further data processing
- ...

Foreseen evolution based on experience

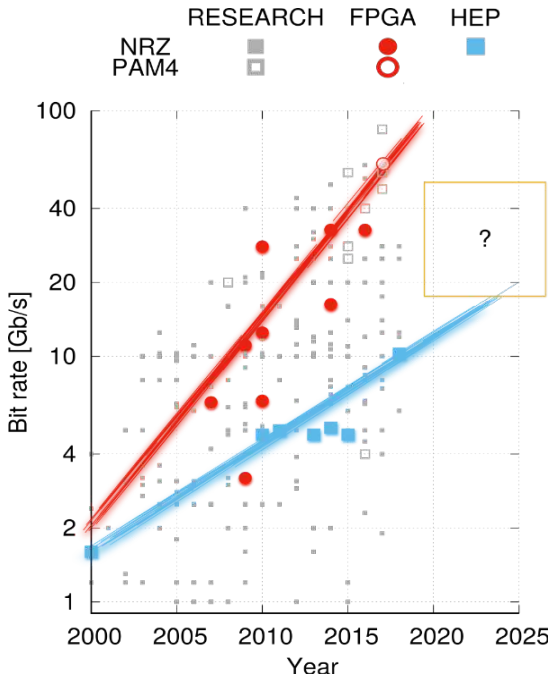
Data center oriented FPGA

Cutting edge FPGA evolution

- Ultra high speed serializers with reduced number of links
  - As of today : 32Gb/s (NRZ) and 116Gb/s (PAM4)
- Well suited for backend DAQ system architecture with latest commercial protocols (PCIe Gen 5, CXL, 400G, 800G)

Front-end evolution

- Increasing link speed of rad-hard serializers takes a huge effort and time
- Is data aggregation feasible to saturate link bandwidth and reduce number of links ?



FPGA vs ASIC link speed (Szymon Kulis)

Toward a split between cutting edge FPGA and front-end serializers

- Keeping compatibility of future front-end serializer and cutting edge FPGA is crucial

Clock distribution path

Requirements for clock distribution

- Lower jitter to the front-end for more precise timing measurement $\mathcal{O}$(10)ps
- Fine phase adjustment control for better calibration and stability of system from reset to reset $\mathcal{O}$(100)ps

Generic readout board with clock distribution in mind

- Use of jitter cleaner external PLL to minimize jitter with many clock schemes to accomodate all use case with clean clocks
- Several strategies to achieve phase control using external PLL with phase control ability and DDMTD*

Requires efforts on tests and characterization a realistic test bench to conform with tighter timing requirements

**DDMTD Digital Dual Mixer Time Difference*

Synthesis

PCIe400 : On-going development

- Evolution of PCIe40 to accomodate with higher bandwidth and tighter timing requirement
- Prototypes are expected by end of summer 2023
- Target integration on few sub-detectors for LS3 on LHCb and others ?

PCIe400 is a stepping stone to prepare for future generic readout board

- Test several future features such as White Rabbit clock distribution, higher speed serializers for front-end up to 25Gb/s NRZ, complex data processing
- Explore new path for DAQ architecture with a network interface on-board

Beyond PCIe400

- Aim at doubling at least the output bandwidth to pursue connectivity with data center standards
- Target development for LS4 with adequate technology available at the time and fitting with refined DAQ system architecture