Technical challenges designing a common readout board for LHCb

Topical Workshop on Electronics for Particle Physics
30th September - 4th October 2024 – Glasgow

Julien Langouët *(CPPM) on behalf of the PCIe400 team – IN2P3*
langouet@cppm.in2p3.fr

Outline

Context

- Goals and rationale
- PCIe400 overview
- Routing a high density board

Thermal dissipation

Power Integrity

Signal Integrity

Summary

Goals and rationale

Generic readout board interfacing as many custom links from front-end as possible with modern commercial links for back-end system

- Can be used in several experiment context

Designed for LHCb LS3 enhancement as a stepping stone for the Upgrade II

- Increase x4 the output bandwidth from current readout board PCIe40
- Explore experimental path for new data acquisition topology such as integrating a network interface on-board, or add complex data processing such as tracks primitive reconstruction
- Distribute LHC master clock with a reproductive phase determinism $\mathcal{O}(10)$ ps pk-pk

PCIe400 overview

Designed around latest largest Altera FPGA Agilex 7 M-series

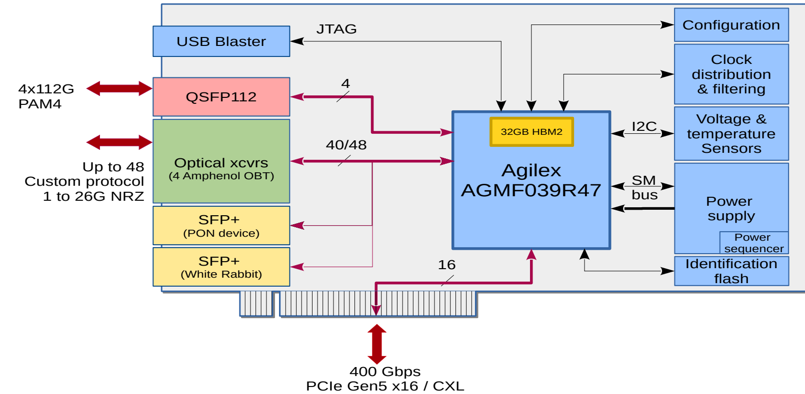
- 4 Million of logic elements and <1 GHz internal frequency
- 32 GB integrated High Bandwidth Memory (HBM)
- Up to 5.2 Tbps exchange between Fabric and HBM
- Arm Cortex 4 cores co-processor

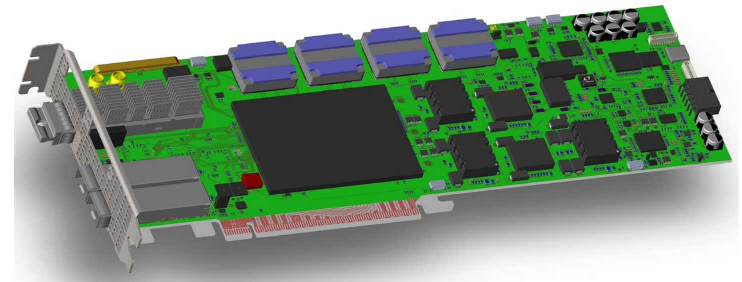
High bandwidth I/O

- Up to 48 bidirectional links with front-end at up to 25 Gbps
- PCIe Gen 5 x16 with 400 Gbps bandwidth – also CXL capable
- 4x bidirectional 112 Gbps for network interface
- 2 SFP+ for Time Fast Control system or White Rabbit

Time distribution

- High precision PLL with <100 fs jitter intrinsic





3D rendered view of PCIe400

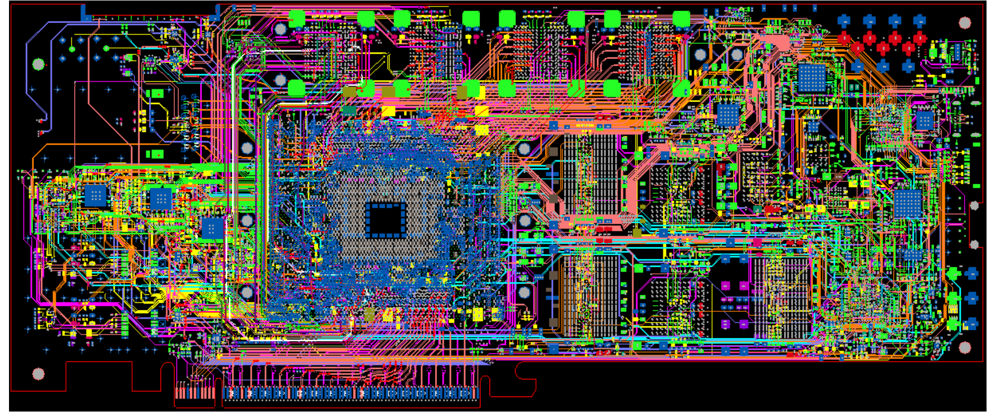
Routing a high density board

PCIe GEN 5 Add-in card double width 268 mm

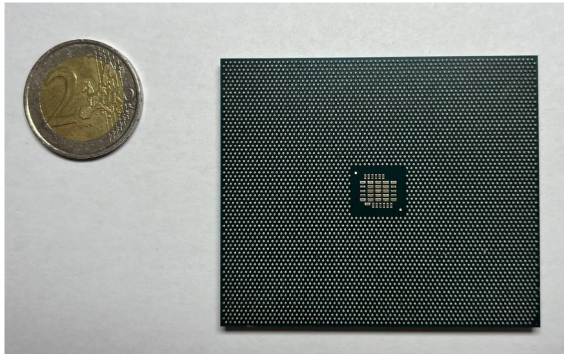
- 2500 components on-board
- 10 000 connections

4500 pins FPGA

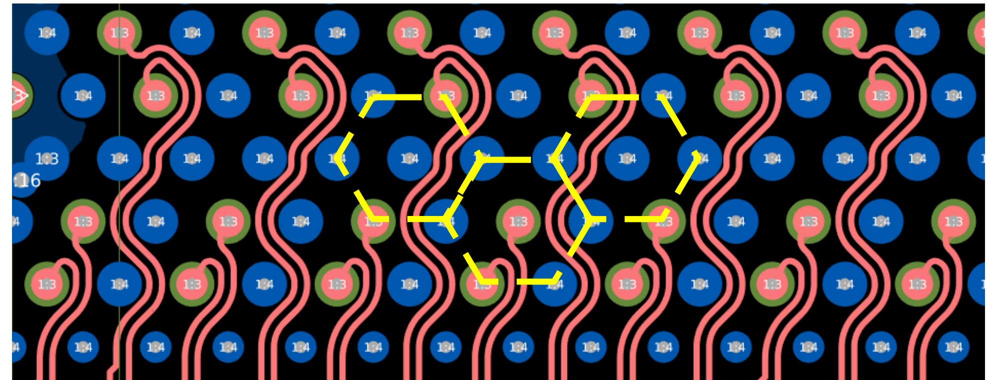
- 0.9 mm pitch BGA in hexadecimal structure



PCIe400 Layout illustration



FPGA BGA pin photo



Hexadecimal ball grid array FPGA

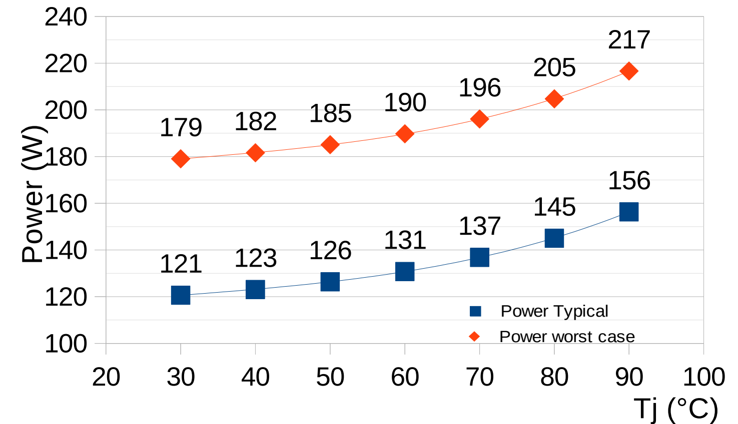
Thermal dissipation

- Power estimation
- Opto-electronic CFD simulation results
- FPGA CFD simulation results
- Prototype heat-sink

Power estimation

FPGA total dissipated power (TDP) estimation at early stage

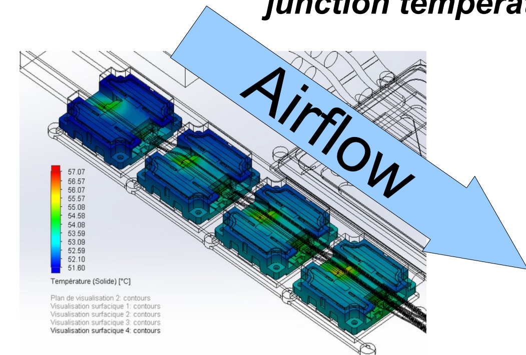
- Based on PCIe40 resource usage in LHCb scaled to Agilex 7 M and applied to its power model
- Estimated between 120W to 230W
- Power consumed by the FPGA exponentially depends on junction temperature → due to leakage current



FPGA power consumption in function of junction temperature

Opto-electronic transceiver

- Estimated as constant: 30 W over 4 modules from datasheet
- Cumulative heat effect due to placement constraint



Opto-electronic transceiver CFD simulation heatmap

Air cooling solution is preferred for its simplicity in terms of infrastructure

Opto-electronic CFD simulation results

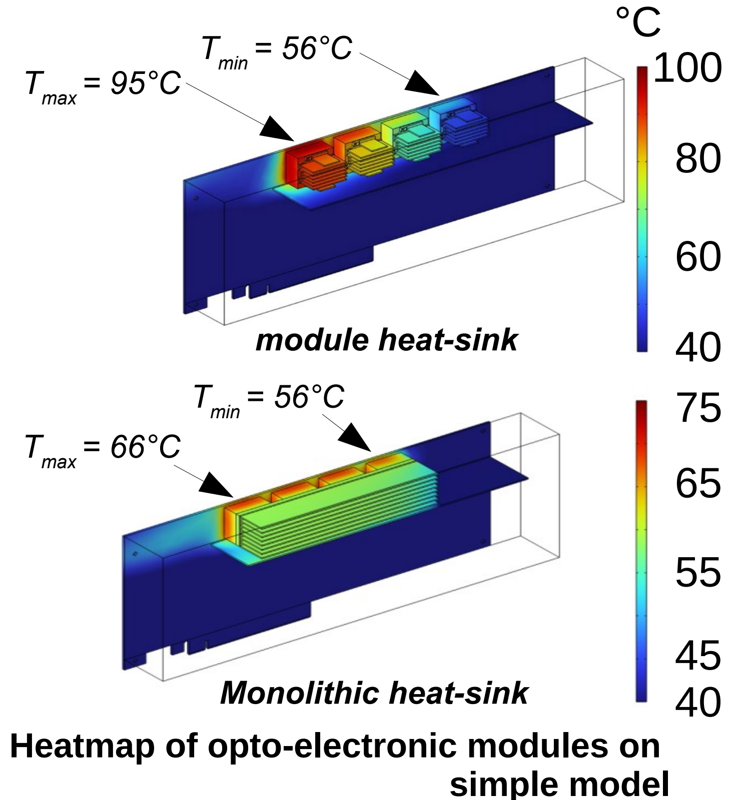
Module heat-sink from manufacturer vs monolithic heat-sink

- Conditions : ambient 38°C, air flow 3 m/s
- Heat-sink aluminum alloy
- 5.7 W per module distributed on the volume

Average temperature dropped by 20% with monolithic heat-sink

- Turbulence in between module heat-sink increase air resistance

Need to design a custom heat-sink for opto-electronic modules



Heatmap of opto-electronic modules on simple model

FPGA CFD Simulation results

Placement constraint allow for large heat-sink surface compared to FPGA package size

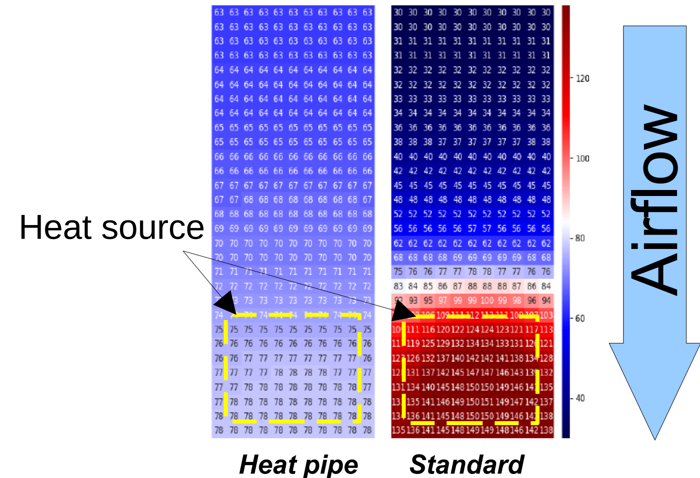
- Need to drain heat away from FPGA to the full heat-sink surface → embed heat pipe in heat-sink base

From CFD simulation, FPGA temperature is extracted

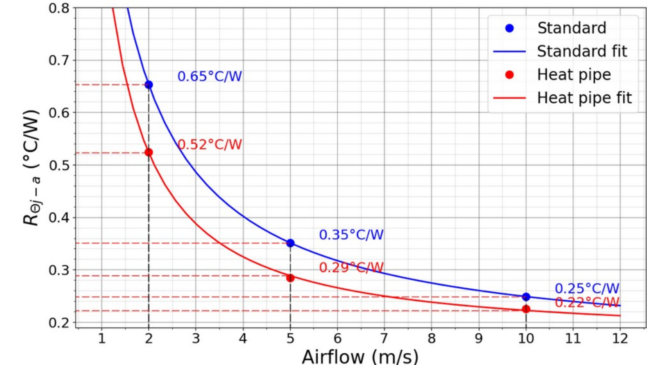
- Thermal resistance $R_{\Theta\ j-a} = \frac{T_j - T_a}{P}$

Heat-sink with heat pipe embedded show ~20% better performance

- With Ta = 38°C, V= 5m/s conditions,
  the standard heat-sink can dissipate 135 W
  the heat-pipe heat-sink can dissipate 165 W
  and maintain the FPGA junction temperature at 85°C



Heat spread on heatsink base illustration (source heatscape)

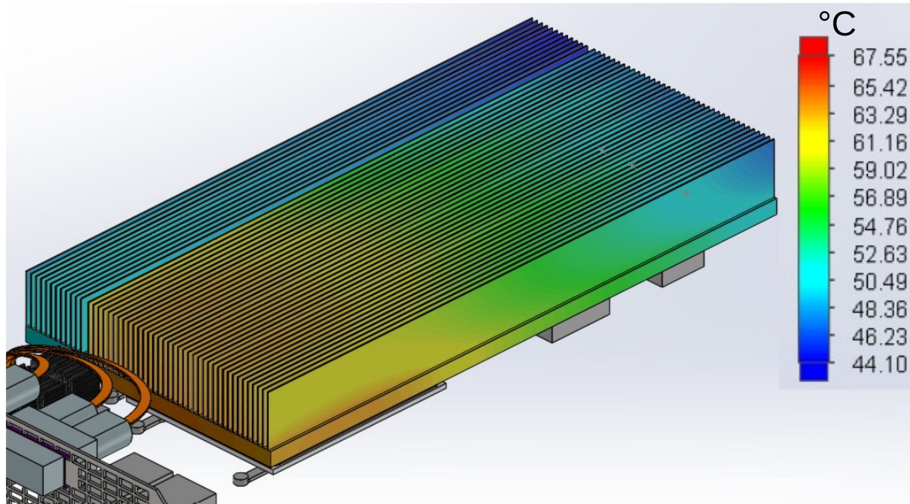


FPGA heat-sink thermal resistance comparison from CFD simulation results

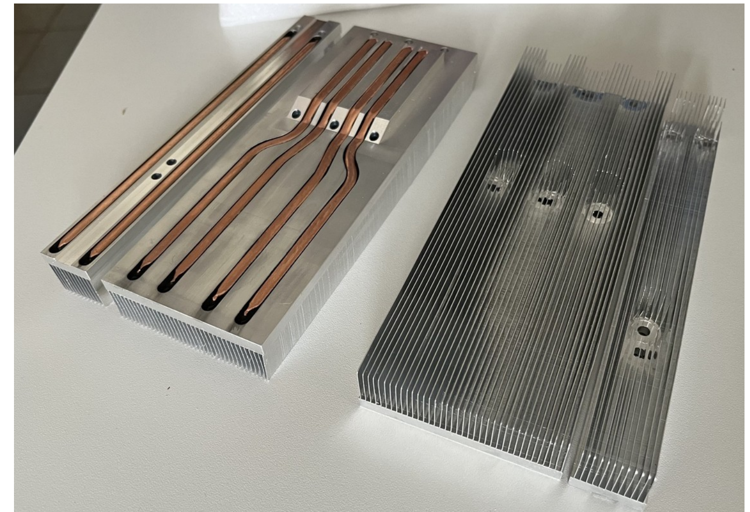
Prototype heat-sinks

Heat-sink final design and simulations outsourced

- Aluminum skived fins with heat pipe embedded



CFD simulation full simulation heatmap
38°C ambient temperature, 5 m/s airflow



Prototype heat-sinks

Power integrity

- Power distribution network
- PCB thermal loss

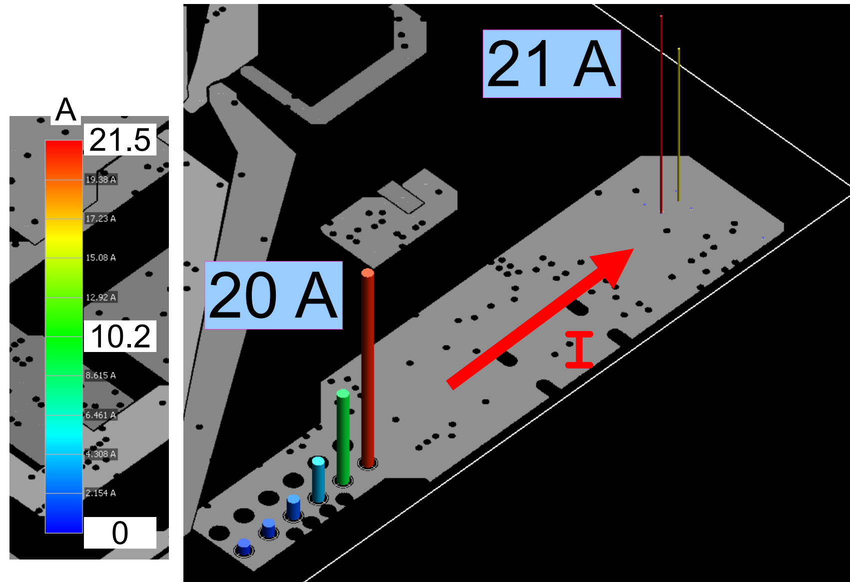
Power distribution network

Power tree

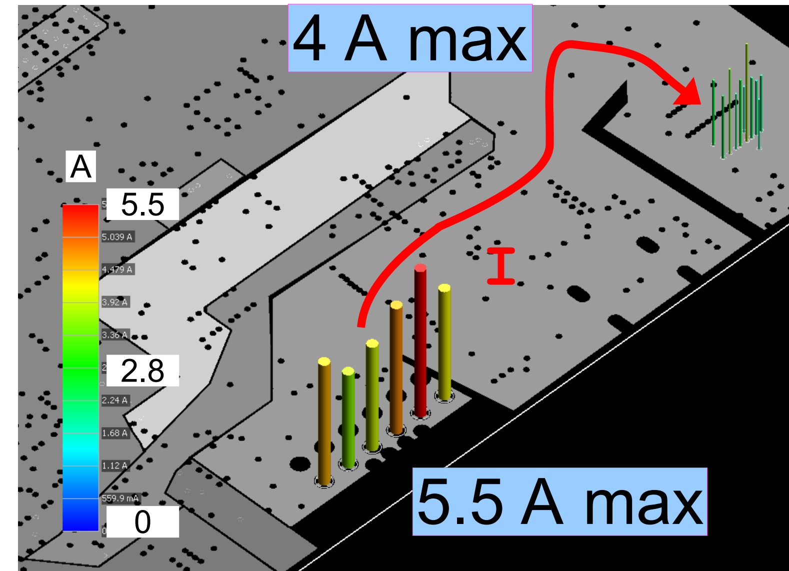
- 22 power rails on board from 0.8 to 5 V
- Maximum estimated current >100 A
- Voltage accuracy down to 0.5 % required

Systematic voltage drop analysis on each power rails at absolute maximum current rate

- Minimize heat loss in power planes
- Distribute current over vias
- Mitigate current bottlenecks



Current in vias
Original power plane design



Current in vias
updated power plane design

PCB thermal loss

Thermal simulation in PCB with Cadence Celsius Thermal Solver

- Based on Altera's reference design DK-DEV-AGI027RES
- VCC core is distributed on Top and L9 layers
- Ambient temperature 25 °C and 2 m/s airflow
- FPGA sinks 200 A
- Change stack-up with 35 or 70 µm copper planes

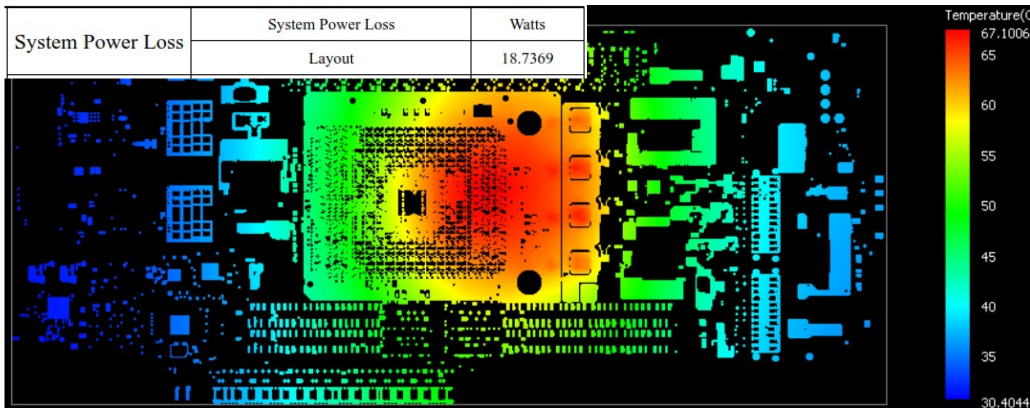
70 µm power planes are used for final stack-up

- Thinner zone near edge connector to respect PCIe standard

Power dissipation within power plane

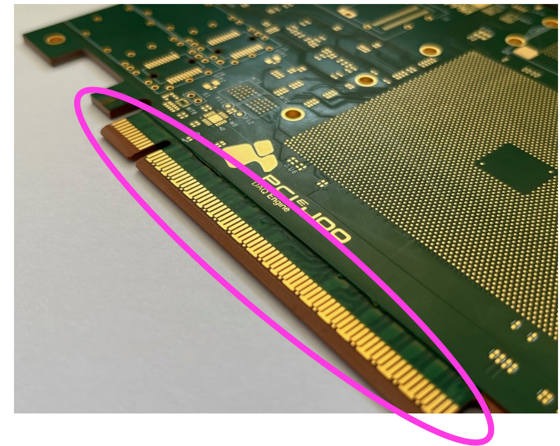
Top & L9 thickness	70 µm	35 µm
TOTAL power loss	11.2 W	18.7 W
PCB T° rise	23 °C	42 °C

PCIe400 mechanical sample PCB



System Power Loss	System Power Loss	Watts
	Layout	18.7369

PCIe400 mechanical sample PCB



Signal integrity

- 2D vs 3D S-parameter extraction
- Placing ports in Clarity 3D
- Opto electronic SMD pads ac-coupling
- FPGA breakout coupling
- Eye diagram at 112 Gbps

2D vs 3D S-parameter extraction tool

Comparison of insertion loss on Altera's reference design

- Cadence Power SI does not consider 3D geometry
- vs Cadence Clarity 3D

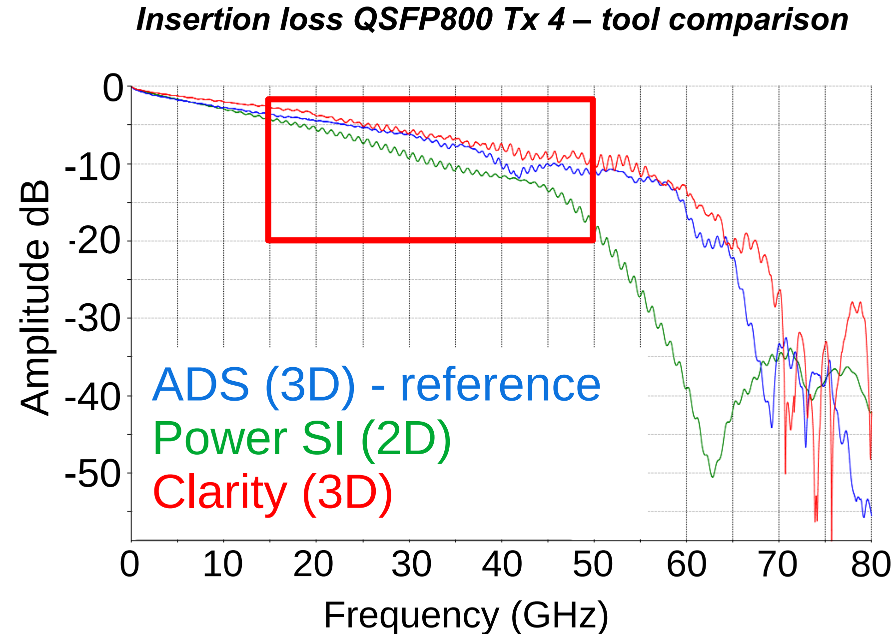
Simulation tools take a lot of computational resources

- S-parameter extraction of a single lane takes ~1 to 6 h on a 64 cores @2.5 GHz machine

136 lanes at up to 112 Gbps PAM4 with 50 GHz bandwidth

- Simulation on most critical of each type of lanes

Insertion loss QSFP800 Tx 4 – tool comparison



Placing ports in Clarity 3D

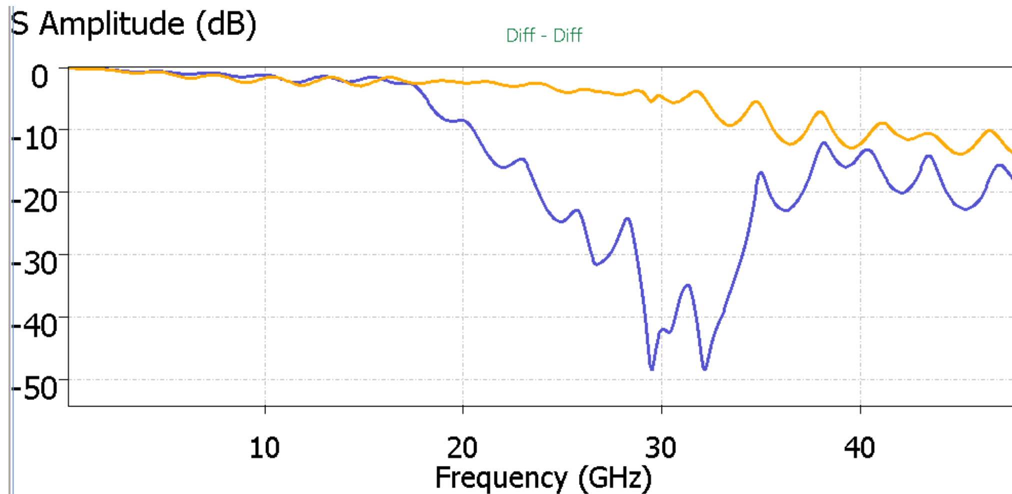
Observed resonance attenuation on PCIe lanes

Common definition is to use pad center as node

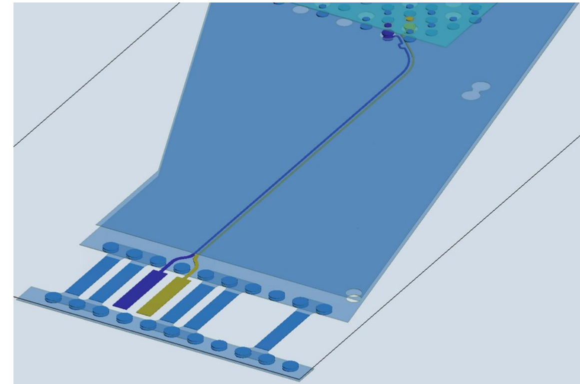
- no impact on resulting physical design
- while Clarity 3D sees a stub

Moving the pad node to the far edge solves the issue

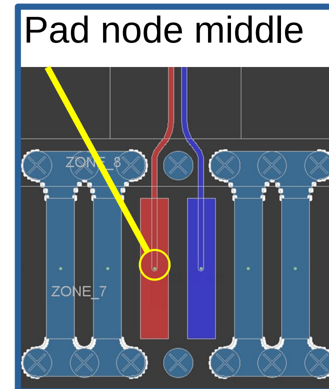
Insertion loss SDD21 – pad node placement comparison

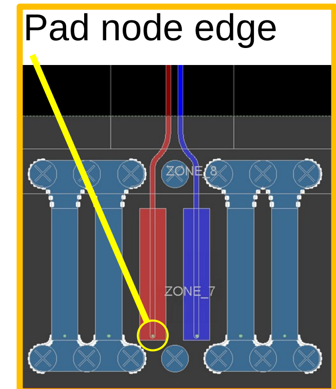


3D view of a PCIe lane



Focus on gold finger node location





SMD breakout optimization pad ac-coupling

Pads from SMD QSFP connector creates a parasitic capacitance with planes underneath

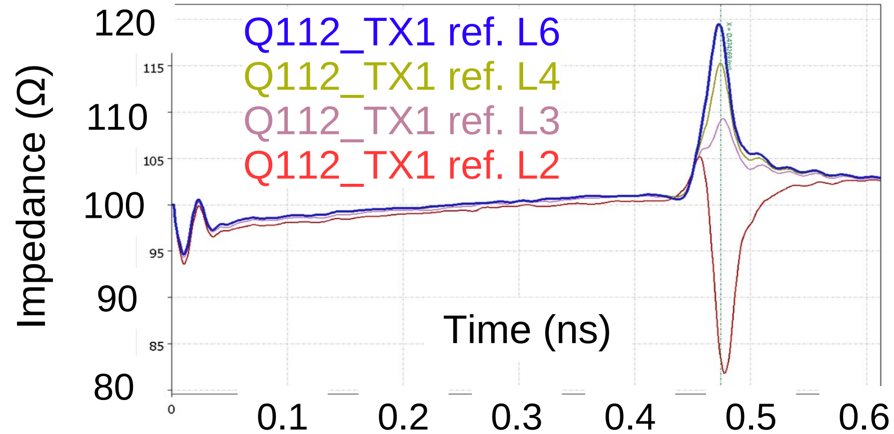
- capacitance is inversely proportional to reference plane distance
- $C = \epsilon_0 \epsilon_r \frac{A}{h}$

  ϵ_0 *dielectric vacum constant*

  ϵ_r *PCB dielectric relatice constant*

  A *pad surface*

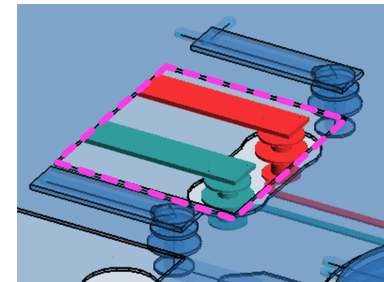
  h distance between microstrip and reference layer



112Gbps TDR response at Tr = 12 ps - impedance mismatch depending on distance to reference plane

Sweeping cut-outs across layers

- Find best balance between inductive and capacitive effect
- Ideal solution may not be achieved depending on layers thickness
- Anti-pad size also plays a role on ac-coupling



Example of connector pad referenced to L3

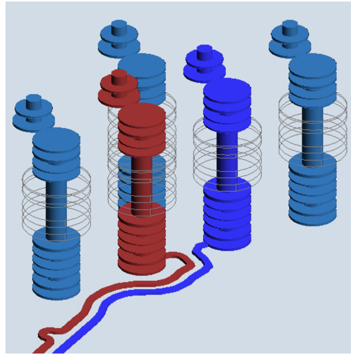
FPGA breakout optimization

PCIe Tx are on bottom layer

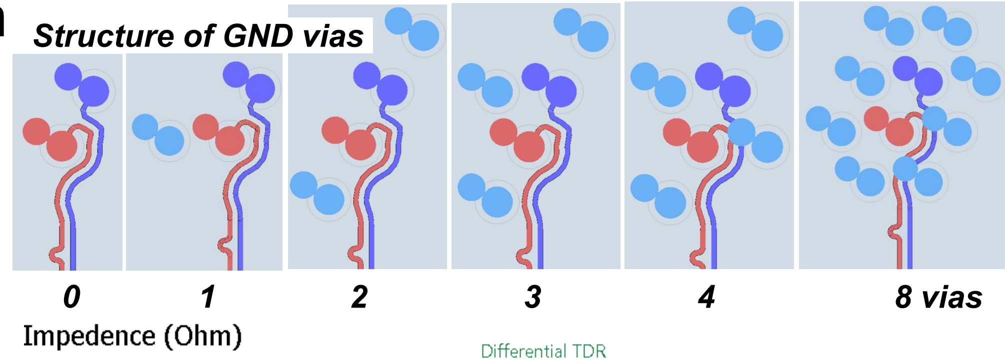
- µvia in pad + buried via staggered in a dense zone
- → can not shield vias with GND vias all around

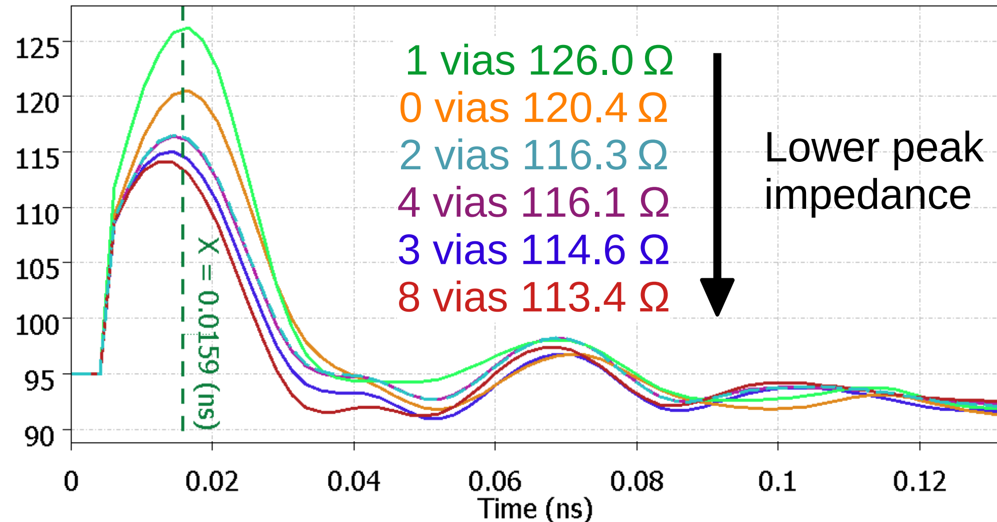
Sweep on number of ground return vias

- Symmetrical structure is important
- 3 vias structure performs better than 4 vias structure are more



3 vias structure – 3D view





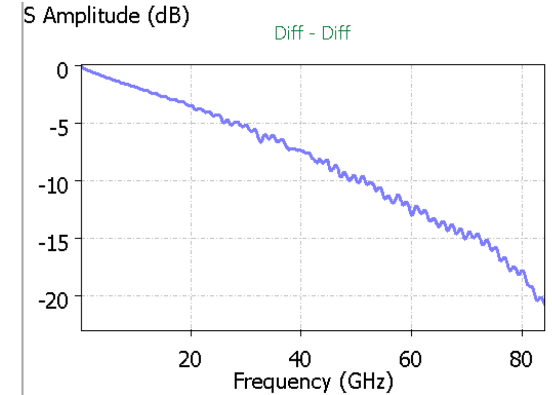
PCIe FPGA breakout TDR response to Tr=15ps, 95Ω

Eye diagram results

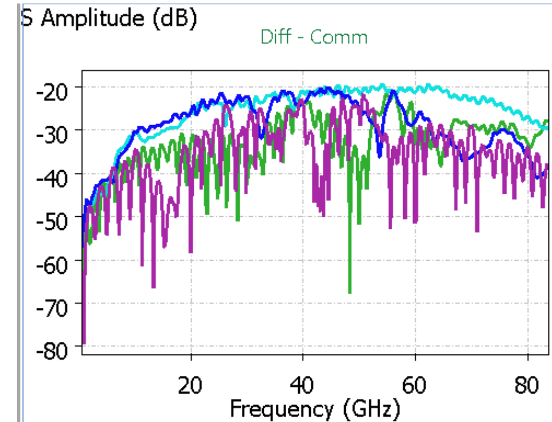
From S-parameter extraction, Intel Advanced Link Analyzer can compute the eye diagram

- Intel expect working lane as long as CDR eye is opened

Rx CDR eye diagram FHT transceiver @106.25 Gbps, UI = 17.86 ps with Intel Advanced Link Analyzer



Insertion loss SDD21



Differential reflection SDC

Summary

PCIe400 is a R&D development pursued by IN2P3 and LHCb Online group

- It embeds modern serial interface such as PCIe GEN5, 4x100 Gbps and up to 48 bidirectional links at 25 Gbps
- Baseline solution for LHCb future upgrade, but it is a generic and versatile module suitable for other application

Design process got into meticulous simulation work for thermal dissipation, power and signal integrity

- CFD simulation resulted in designing custom heat-sink with heat-pipes capable of dissipating 195 W with 5 m/s airflow
- Systematic power distribution network verification with layout iteration to optimize power plane and via current
- Learn on benefits and pitfalls of S-parameter 3D extraction tool: Clarity 3D necessary for > 25 Gbps serial links
- As s-parameter extraction is time consuming, could not test each and every lanes
  → use of layout structure patterns to ensure consistent design on the 136 serial lanes

First prototypes are being manufactured, expected by end of November 2024

Back-up

Power Estimation for FPGA

Static power depends on :

- Resource activation
- Junction Temperature

Dynamic power depends on :

- Resource occupancy
- Frequency
- Toggle rate

Definitive firmware not available for hardware design

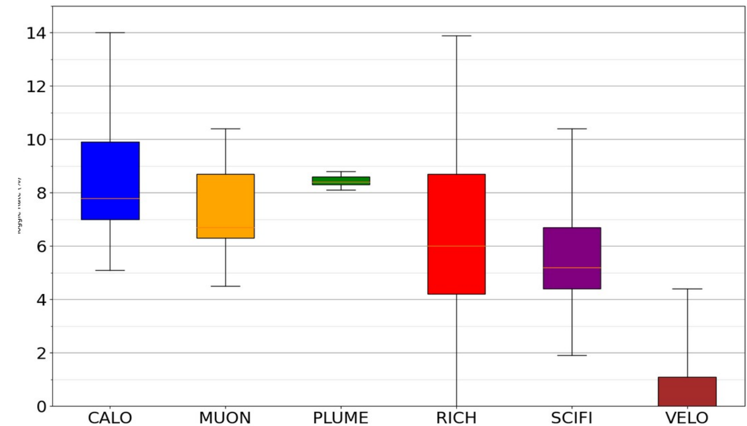
- Use of Intel's tool with Agilex power model
- Use of TELL40 firmware statistics

Toggle rate is a predominant factor and impacts the power supply decoupling scheme

- Based on TELL40 FPGA core current measures
- compilation report, the toggle rate is <12.5 %

- 12.5 % toggle rate
- 640 MHz
- Typical case:
  - 60% logic
  - 80% RAM
  - 48 links 10G
- Worst case:
  - 80% logic
  - 100% RAM
  - 40 links 25G + 400GbE

Resource usage considered



Processed global toggle rate (%) by subdetector (july 2022)