# PCIe40 Design

**J.P. Cachemiche**, P.-Y. Duval, F. Hachon,
M. Jevaud, R. Le Gac, F. Rethore
**Centre de Physique des Particules de Marseille**

## Outline

- PCIe40 Features
- Dimensionning
- Status

PCIe40
DAQ Engine

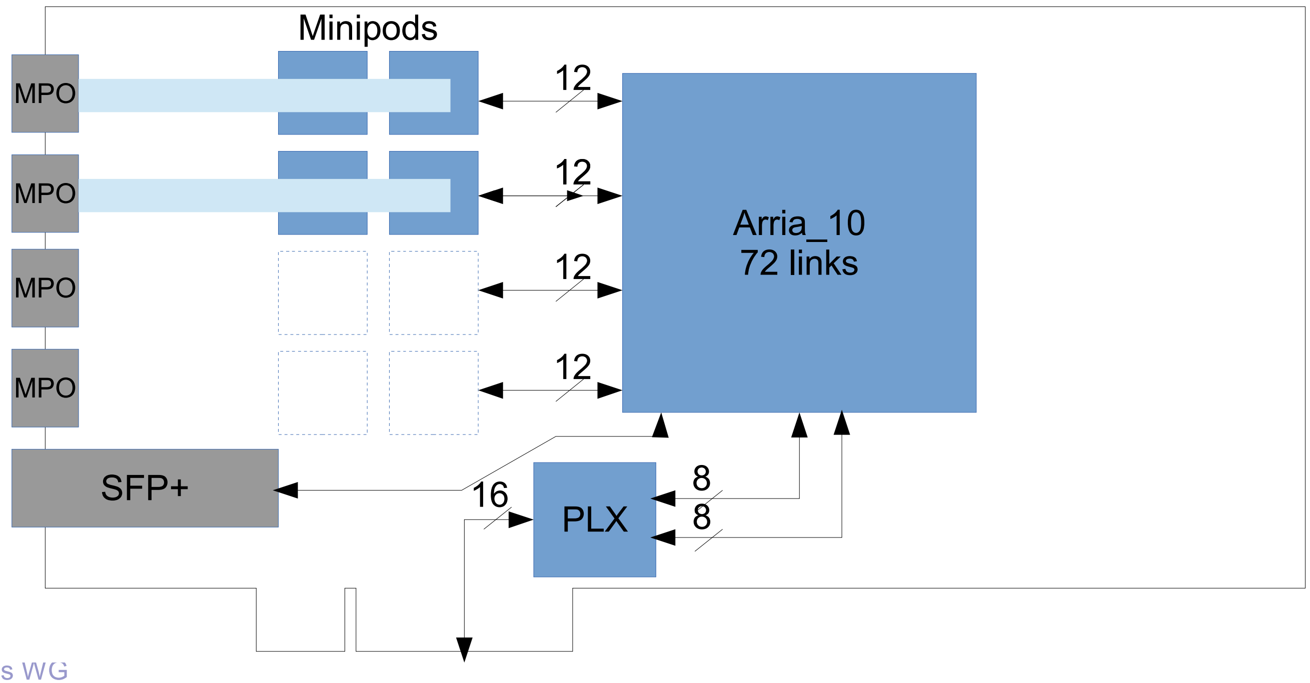
# PCIe40 board features

## Nominal configuration:

- 1 bidir link for TFC
- 24 GBT inputs → limited by PCIe output bandwidth
  - PCIe GEN3 x16 = 110 Gbits/s
  - 24 GBT wide bus = 107 Gbits/s

## Extended configuration

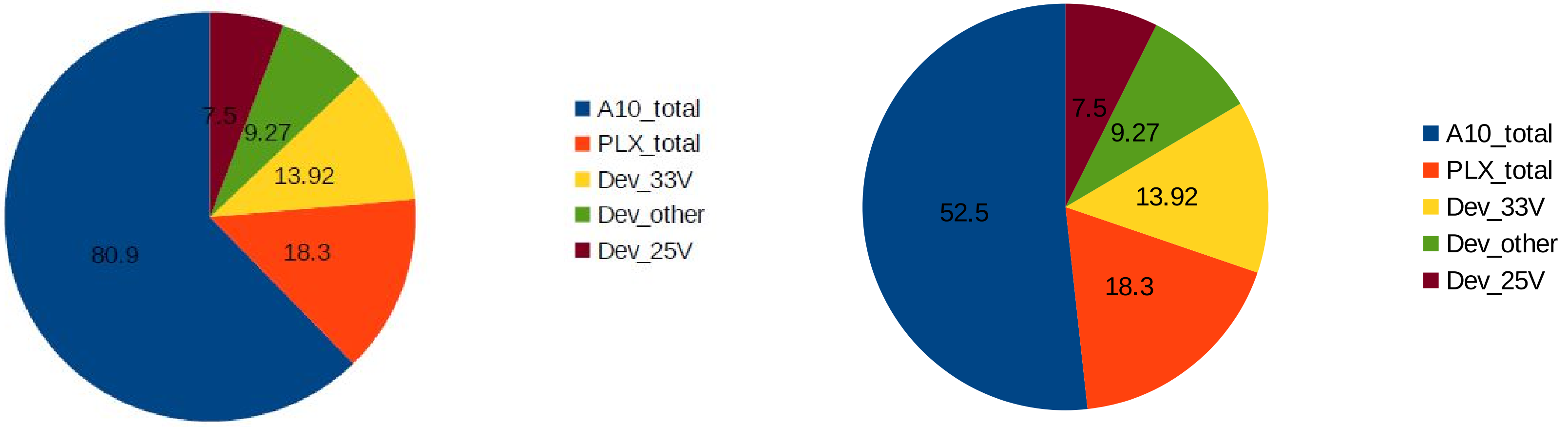
- Up to 48 bidir links available on board for low luminosity sub detectors → decrease the costs



# Power consumption

## Estimation

- Dimensionned for use of 100 % logic cells, 50 % toggle rate, 250 MHz operation, ES chip
  - **157 W** for engineering sample:
  - Derating of 2/2.5 on some features
  - Hopefully much less for production chip (~123 W)



**Engineering sample**

**Production chip**

**Power consumption repartition in W**

(20 % more must be added for total consumption more to take into account DCDC efficiency)

# Layout

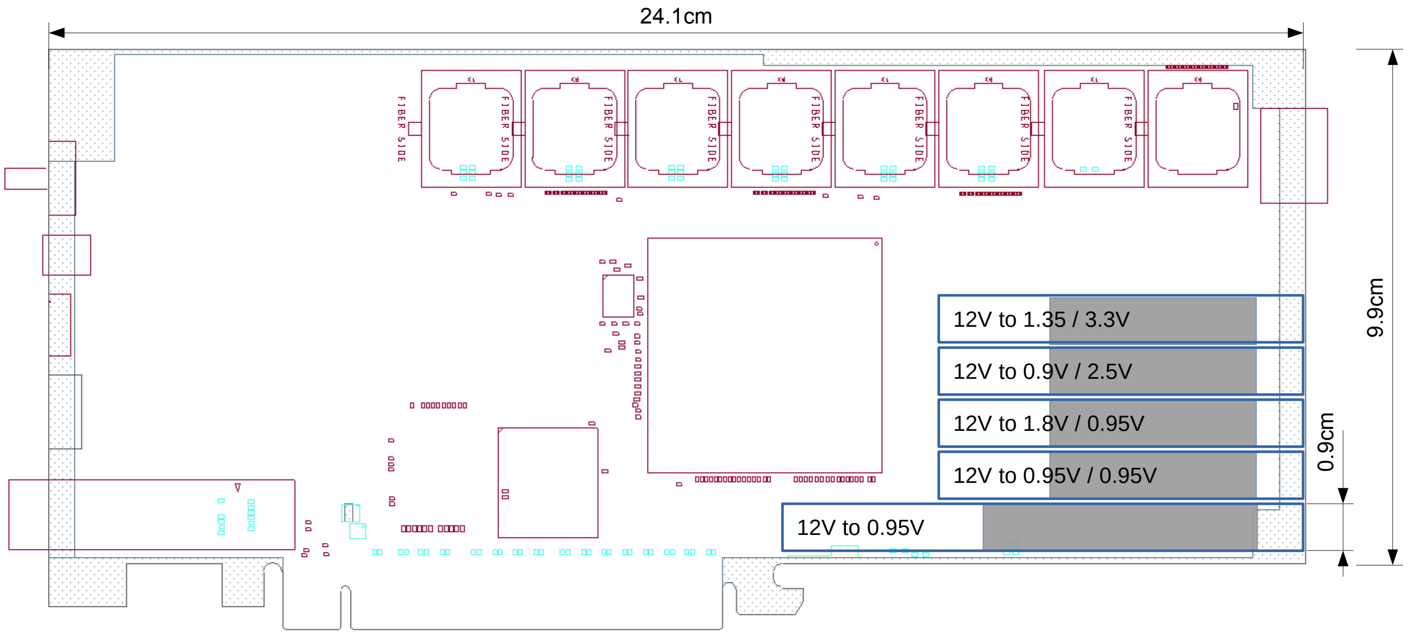
## Layout crisis !

- Power supplies take much place
- Initial estimations based on a 312 mm board and most servers mechanically dimensionned for a 241.3 mm size
- Use of micropods instead of minipods not sufficient to recover enough place

## Solution

- Migrate most of power supplies on vertical mezzanines

# Mezzanine implementation



12V to 1.35 / 3.3V

12V to 0.9V / 2.5V

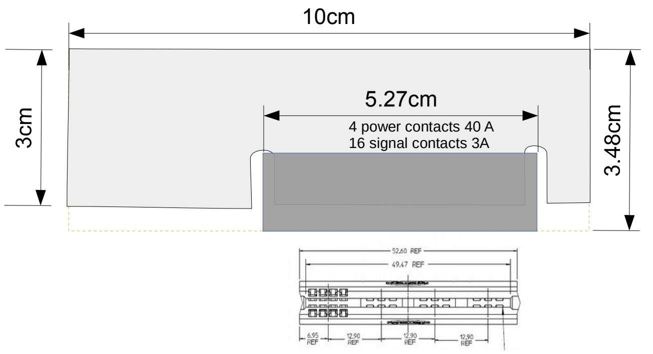
12V to 1.8V / 0.95V

12V to 0.95V / 0.95V

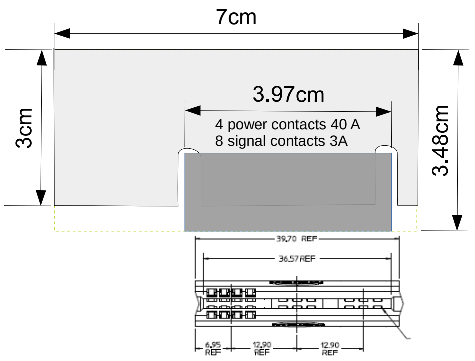
12V to 0.95V

Mating to gold finger card edge - mixed signal segments

Mating to bus bar - power only segments



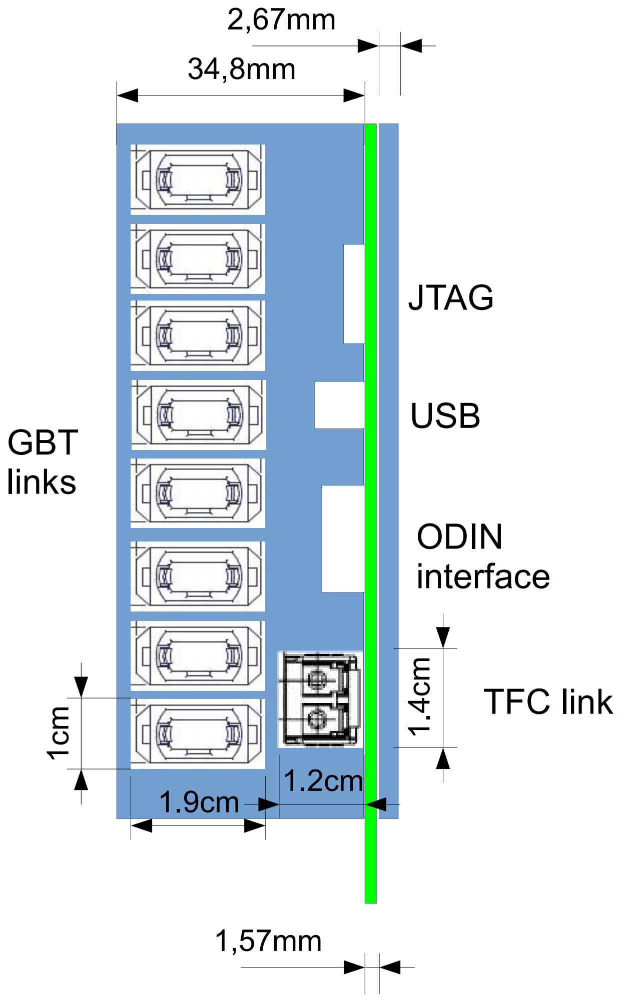
4 power contacts 40 A
16 signal contacts 3A

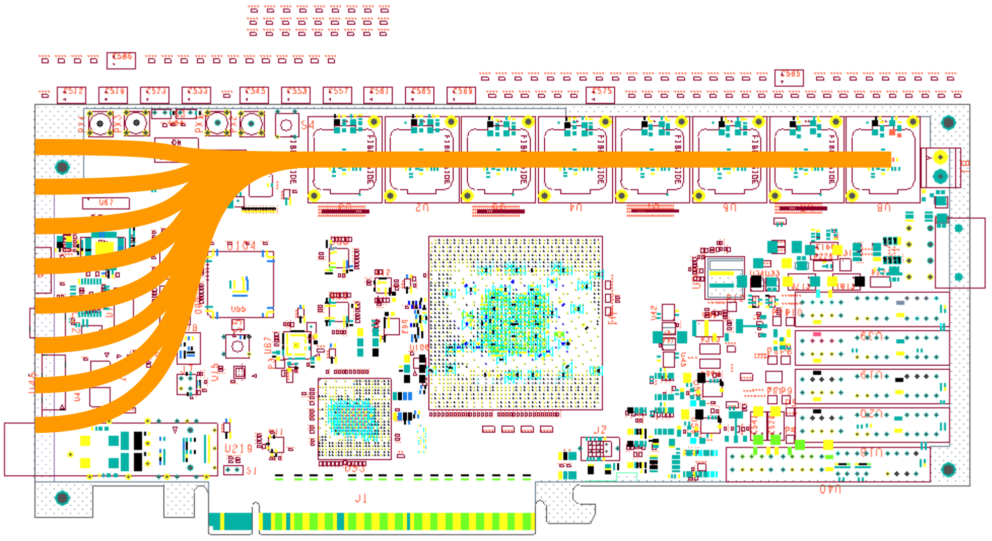


4 power contacts 40 A
8 signal contacts 3A

# Face plate

2 slots size  to let room for heatsink or fan
… but also to place MPO connectors on face plate

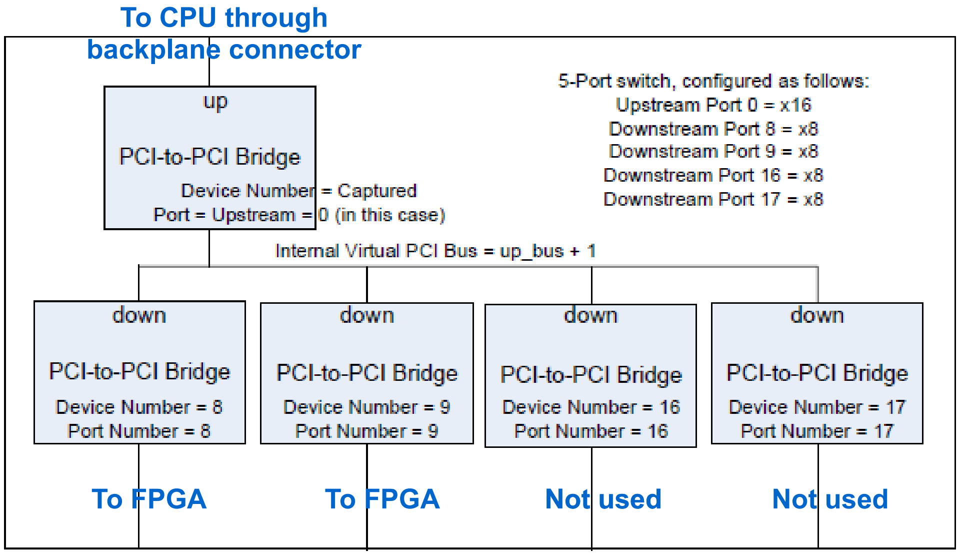




# PCIe interface

## PLX 8747 chip for converting two x8 PCIE into one x16 PCIe

- Could be latter removed if board used in a CPU with Haswell architecture
- Does not allow to test it in the meantime because supports only one upstream port
- Possible change for a 8748 able to address both configurations
- To many modifications required
  - ➨ Decision not to change anything in the first version of the prototype



To CPU through backplane connector

5-Port switch, configured as follows:
Upstream Port 0 = x16
Downstream Port 8 = x8
Downstream Port 9 = x8
Downstream Port 16 = x8
Downstream Port 17 = x8

up
PCI-to-PCI Bridge
Device Number = Captured
Port = Upstream = 0 (in this case)

Internal Virtual PCI Bus = up_bus + 1

down
PCI-to-PCI Bridge
Device Number = 8
Port Number = 8

down
PCI-to-PCI Bridge
Device Number = 9
Port Number = 9

down
PCI-to-PCI Bridge
Device Number = 16
Port Number = 16

down
PCI-to-PCI Bridge
Device Number = 17
Port Number = 17

To FPGA

To FPGA

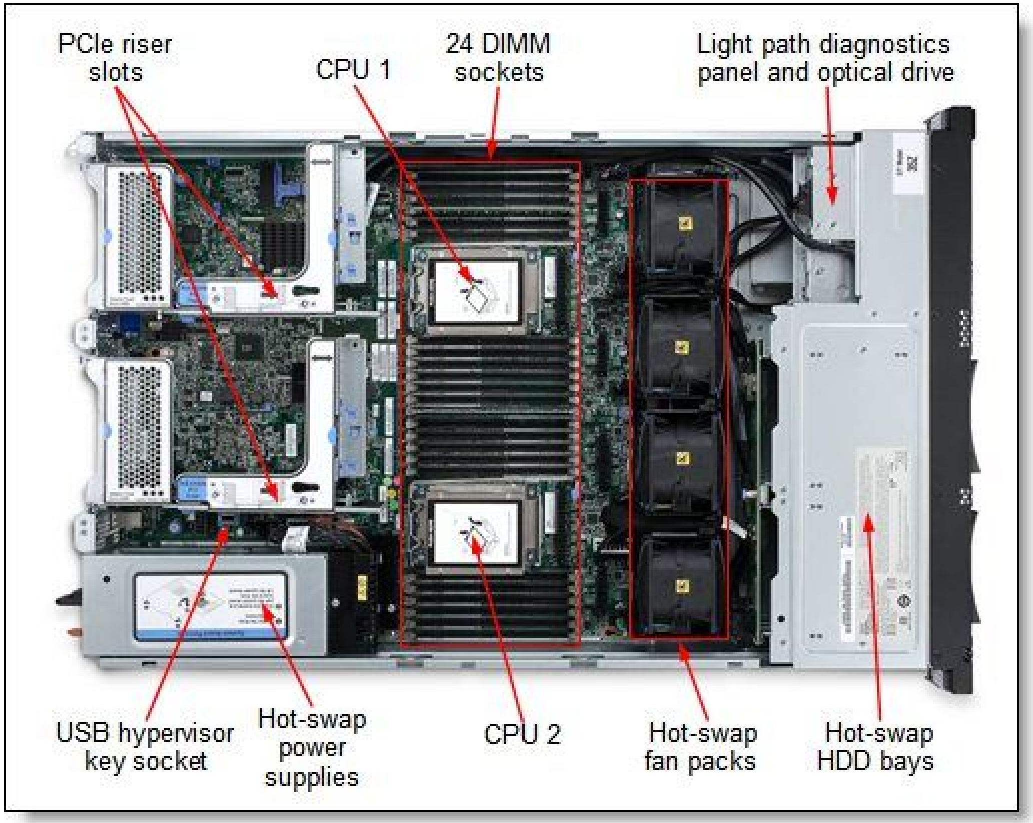
Not used

Not used

# Environment

## Cooling ?

- Air cooled ?
  - Direction ?
  - Strenght ?
  - Ambiant temperature if CPU cooled with same flow?
-

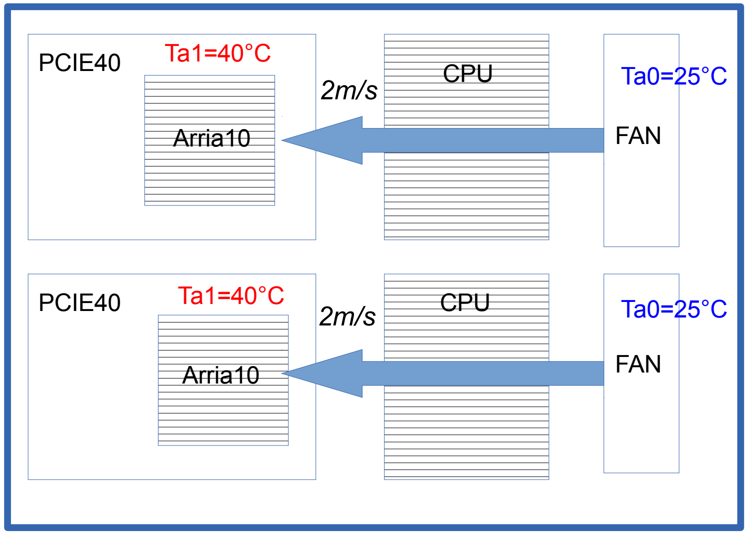


# Cooling dimensionning

## Hypothesis

- Maximum height standard heat-sink same footprint as FPGA
- Use of production chip
- Average clock 250 MHz
- Average toggle rate 50 %
- Ambiant = 40°C
- Air flow = 2m/s
  - ➨ To be checked or specified on chosen server

1U – BLADE - Server



# Precautions taken

## Understanding cooling

- Temperature monitoring in every corner of the board (14 measured points)

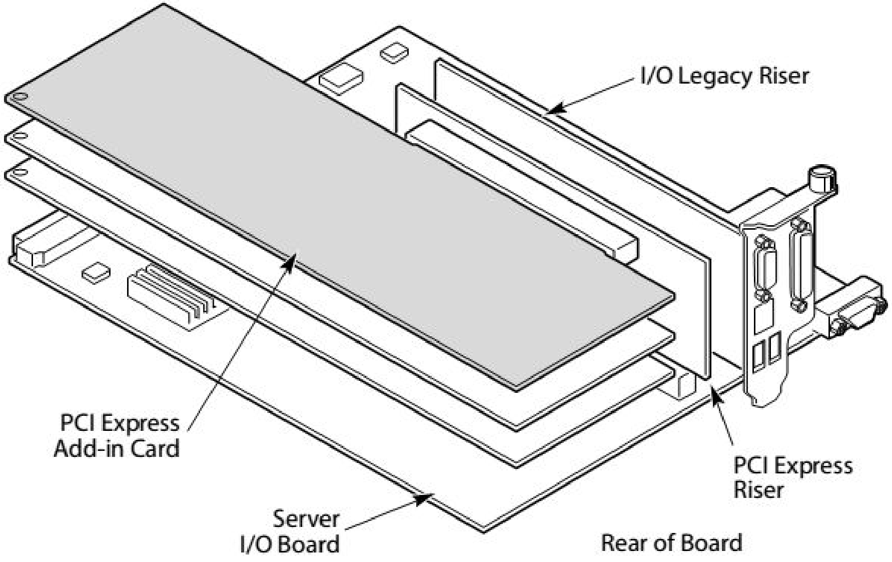
## Understanding power consumption

- Current of each voltage measured (10 measured points)
- Voltages measured (6 measurements)

# Environment

## CPU box form factor

- Mechanical implementation of a PCIe board in a CPU blade makes use of a riser card



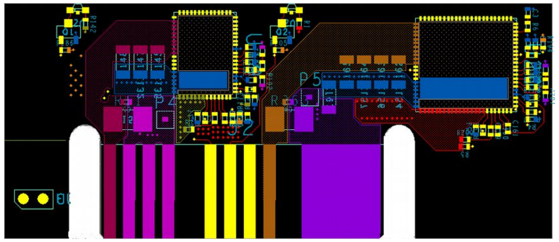
## Can we stay 1U ?

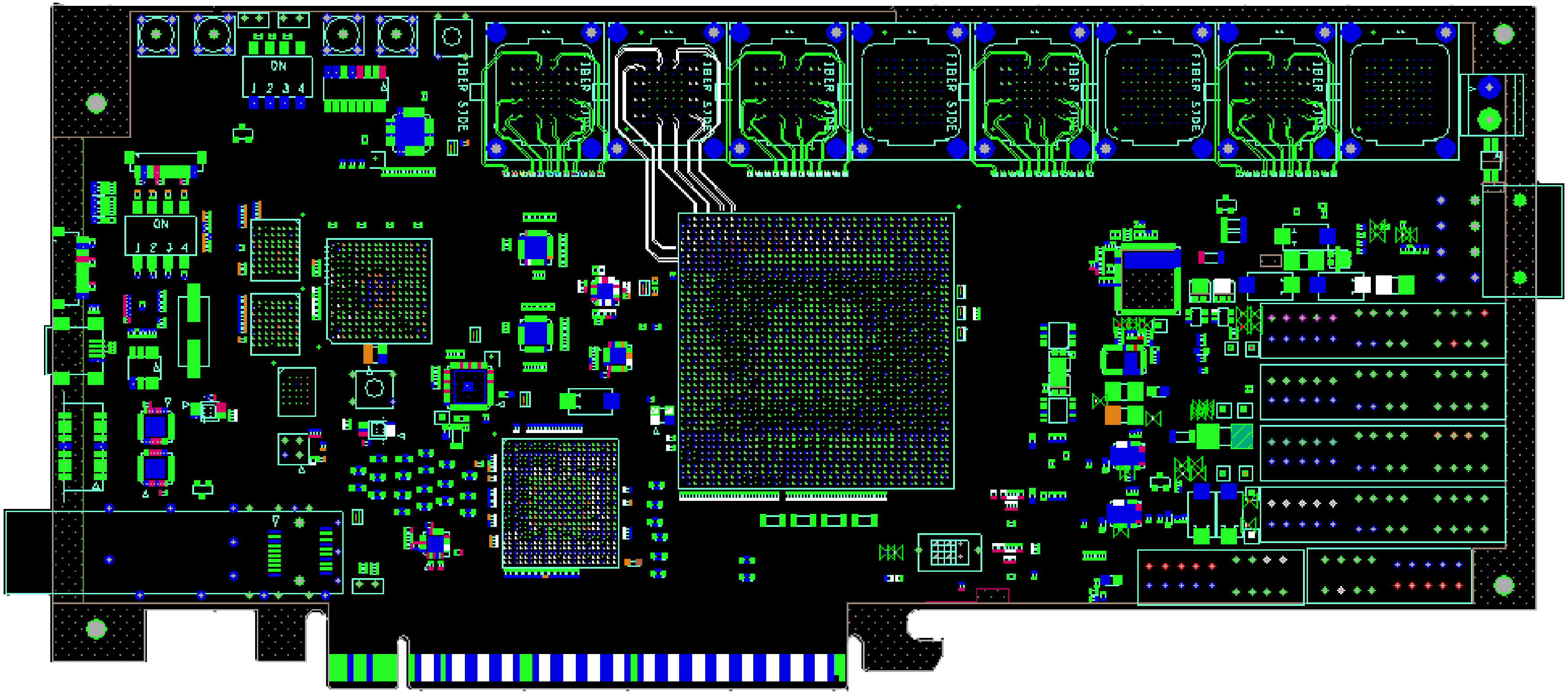
- 2 slot board = 39.1 mm
- 1U = 44 mm
- 4.8 mm remaining for mother board

→ Thin margin

# Status

Routing on-going





# Approximative schedule

**Some delay because of layout crisis due to the reduced size of the board**

| Nominal scenario | | | Dec | | | | Jan | | | | Fev | | | | March | | | | April | | | | May | | | | June | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Routing | | | ▨ | ▨ | ▨ | 🎄 | ▨ | ▨ | | | | | | | | | | | | | | | | | | | | | | |
| Manufacturing 1st prototype | | | | | | | | | ▨ | ▨ | ▨ | ▨ | ▨ | | | | | | | | | | | | | | | | | |
| Debug | | | | | | | | | | | | | | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ | ▨ |
| Manufacturing 2nd prototype | | | | | | | | | | | | | | | | | | | | | | | ▨ | ▨ | ▨ | | | | | |

# Conclusion

## Development is ongoing

- Few difficulties of layout because of reduced form factor, now solved
- Routing of the board and power mezzanines ongoing
- Prototype will not address Haswell architecture in this first version
- First prototype expected end February 2015

## Joint cooling study to be done by Online/CPPM

- Environment has to be understood for existing servers
  - Available space for the card
  - Cooling characteristics
- If not appropriate, see how to customize it