DE LA RECHERCHE À L'INDUSTRIE

POSSIBLE CONTRIBUTIONS OF DEDIP IN ELECTRONICS TO HYPER KAMIOKANDE

D. Calvet,
CEA Paris-Saclay

Saclay, 24 November 2020

www.cea.fr

POSSIBLE PERIMETERS OF CONTRIBUTION

A new QtC ASIC

- Development of HKROC led by Omega group

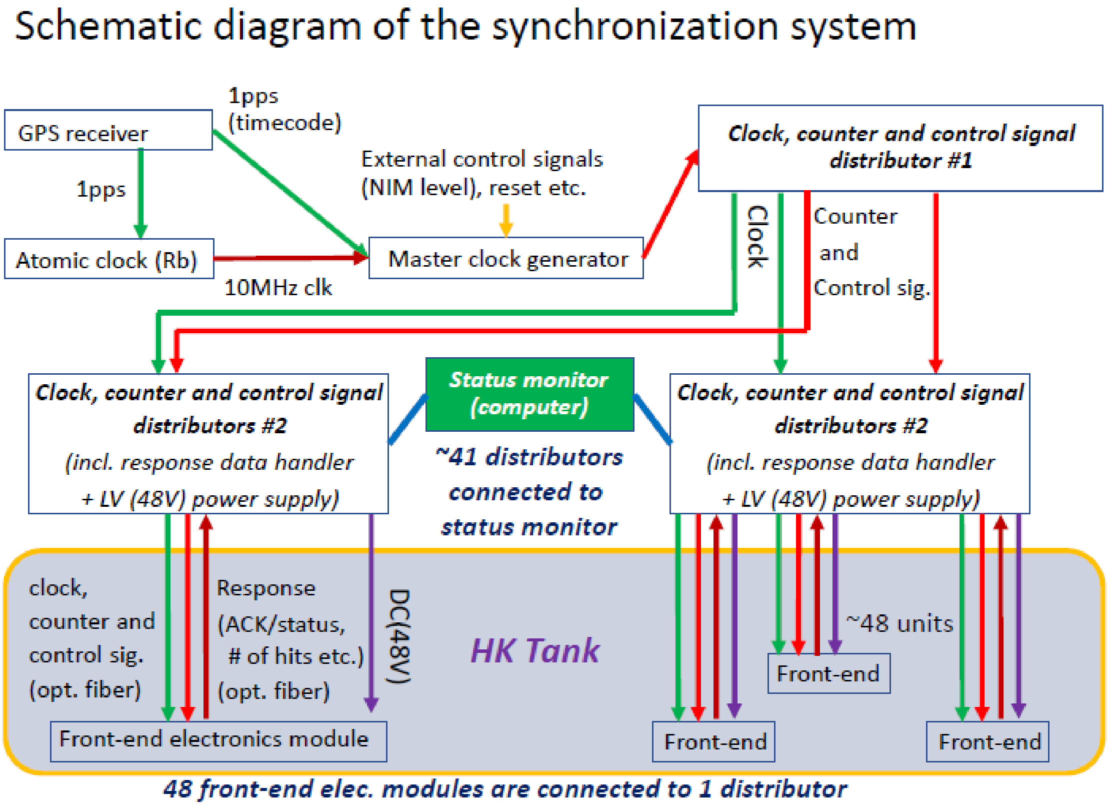
  → Currently no involvement of DEDIP. Some discussions on-going for a possible contribution (see e.g. CMS HGROC)

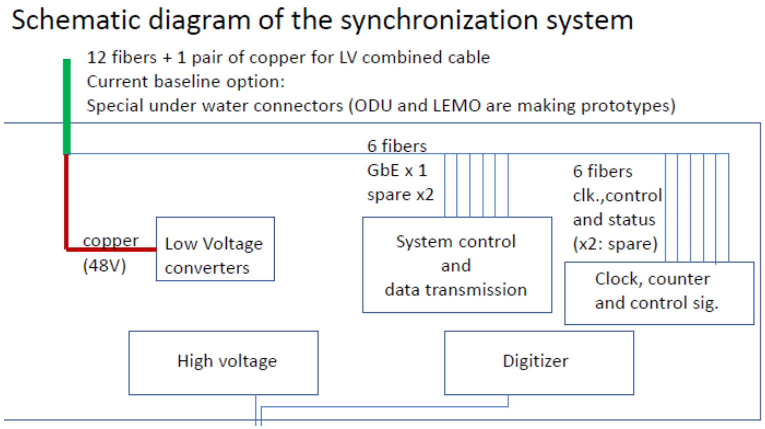
Clock Distribution System

- On-going effort led by Lpnhe with contributions from INFN and Tokyo University groups

  → Starting discussions on how DEDIP could join this sub-project

SYNCHRONIZATION SYSTEM FOR HYPER K (AS SHOWN MID-2019)


Schematic diagram of the synchronization system
GPS receiver
1pps (timecode)
1pps
Atomic clock (Rb)
10MHz clk
External control signals (NIM level), reset etc.
Master clock generator
Clock, counter and control signal distributor #1
Clock
Counter and Control sig.
Clock, counter and control signal distributors #2 (incl. response data handler + LV (48V) power supply)
Status monitor (computer)
~41 distributors connected to status monitor
Clock, counter and control signal distributors #2 (incl. response data handler + LV (48V) power supply)
clock, counter and control sig. (opt. fiber)
Response (ACK/status, # of hits etc.) (opt. fiber)
DC(48V)
HK Tank
~48 units
Front-end
Front-end electronics module
Front-end
Front-end
48 front-end elec. modules are connected to 1 distributor


Schematic diagram of the synchronization system
12 fibers + 1 pair of copper for LV combined cable
Current baseline option:
Special under water connectors (ODU and LEMO are making prototypes)
6 fibers
GbE x 1
spare x2
6 fibers
clk.,control
and status
(x2: spare)
copper (48V)
Low Voltage converters
System control and data transmission
Clock, counter and control sig.
High voltage
Digitizer

Excerpts of presentation made by Hayato-san (mid-2019)

Concept

- Underwater front-end module (FEM) serves 24 PMs. ~2000 FEMs total max.
- Distinct networks for DAQ and clock distribution (custom protocol or not)
- Clock distribution by 3-stage tree: 1 master x 41 slave distributors x 48 FEMs (for 47232 PMs)
- Optical link per FEM: (1 daq + 1 clock + 1 control + 1 status) x (1 nominal + 1 spare)

CURRENT STUDIES ON CLOCK DISTRIBUTION BY HK GROUPS (LPNHE, INFN, U. TOKYO)

The time distribution system consists of 2 main parts

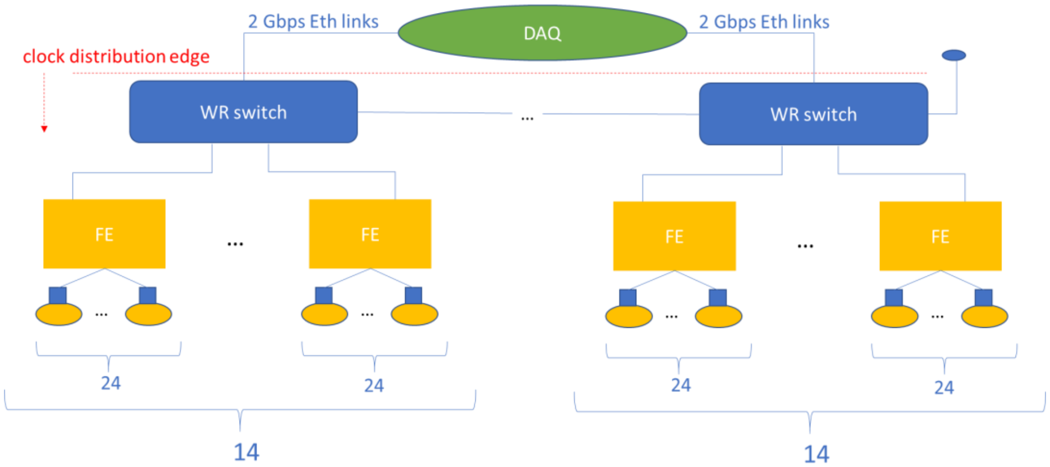
- A system that generates the local time base correlated to the Universal Time Coordinated (UTC).
- A distribution network that delivers the clock to all the front-end nodes and establish a communication link for critical slow control.

Two concepts are subject of our R&D:

  - Direct distribution SK-like.
  - Clock embedded into data (clock and data recovery concept): Custom solution, or White Rabbit

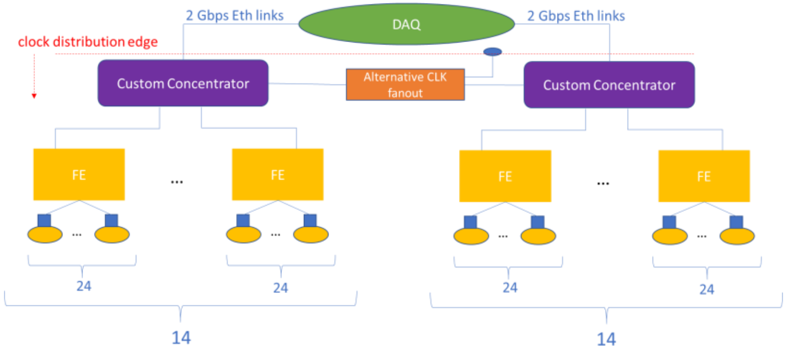
CERN White Rabbit

A possible architecture



Custom Solution - 2

A possible architecture (clock fanout)



Excerpts of presentation made by Stefano Russo at RT2020 conference, Oct. 2020

MY PERSONAL OPINION…

…on Hyper K readout architecture

- Larger number of optical transceivers than needed on FEM. Consider merging some of them
- Separation of clock distribution and DAQ in two distinct networks arguable. Provides independence for development but brings higher cost, complexity, power consumption. Reduces global system availability? (both clock network AND daq network have to be OK)
- 3-stage tree topology good for clock distribution, seems more robust than cascading a large number of switches in series. Current diagrams not so clear to me on DAQ network part.

…on White Rabbit

- Proven technology and will certainly work, but current products on the market not ideally matched: low switch density (18 ports), WR3 switch designed in ~2014 based on Xilinx Virtex 6 almost obsolete; WR4 under development based on a costly Zynq Ultrascale+. Aiming for Ethernet 10G+ but same port count per switch
- The 48.000 PMT scenario would require ~150 switch WR3/4 18-ports switches: expensive, lot of rack space needed, complex cabling, etc.
- Oversized in terms of bandwidth – downstream and also upstream (unless clock distribution network is also used for detector DAQ)
- Unclear if some functionality are not too much, e.g. dynamic clock phase adjustment
- Strong community of users and CERN support but commercial availability depends on two startup companies, Seven Solutions and CreoTech

CAN A CUSTOM SOLUTION BE ADVANTAGEOUS? YES, AND DEDIP CAN BRING TWO INNOVATIONS

A novel method for clock distribution

- Distribute a clock carrying serial data instead of reconstruct a clock obfuscated in serial data
- First results on « Clock Duty Cycle Modulation – CDCM » encouraging
- Potential gains in precision and purity of distributed clock. Simplification of FE receivers: FPGA agnostic, does not require high speed SERDES, only PLL + small logic
- CDCM transmitter reversible to common serial data encoding by firmware change. Receiver optimized for CDCM needs extra hardware for clock/data recovery of usual serial encoding
- CDCM may be used only in back-end to front-end direction (low bandwidth) while ordinary serial data encoding more adequate for links from front-end to back-end (higher bandwidth)

More info? See **https://arxiv.org/abs/2010.14164**

A novel implementation tailored to the specific needs of the application

- Bandwidth adapted to requirements, i.e. asymmetric, not imposed by a standard like 1G or 10G
- Switch core based on an inexpensive commercial FPGA module instead of a high-end device
- No superfluous functionality or unnecessary features. Be application specific, not a universal solution
- Increased density to 48 ports in the same or comparable volume as an 18-port WR3/4 switch
- Baseline design dedicated to clock distribution only; upgradable to serve for main DAQ if the system architecture evolves in this direction

→ ***Danger to avoid: try to build ourselves a bigger, better, faster, cheaper White Rabbit switch***

HYPER KAMIOKANDE DAQ REQUIREMENTS

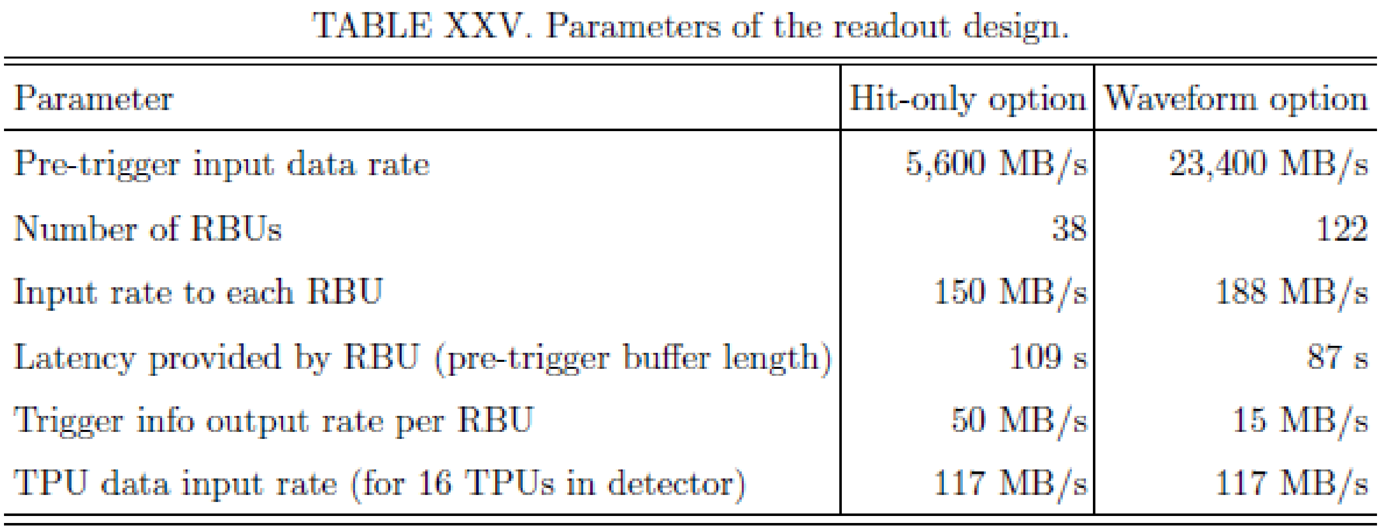


TABLE XXV. Parameters of the readout design.

Parameter	Hit-only option	Waveform option
Pre-trigger input data rate	5,600 MB/s	23,400 MB/s
Number of RBUs	38	122
Input rate to each RBU	150 MB/s	188 MB/s
Latency provided by RBU (pre-trigger buffer length)	109 s	87 s
Trigger info output rate per RBU	50 MB/s	15 MB/s
TPU data input rate (for 16 TPUs in detector)	117 MB/s	117 MB/s

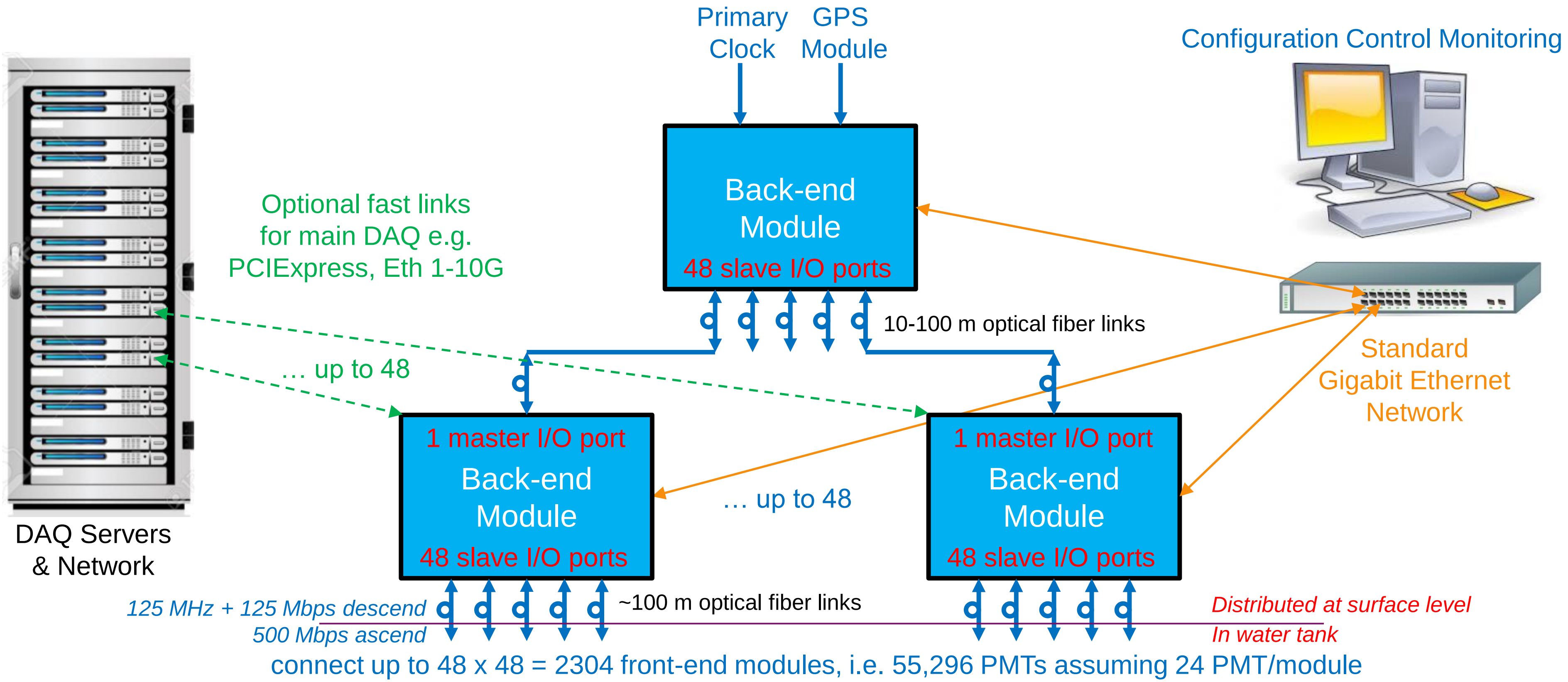
From HK Design Report Nov. 30 2018 (46,700 PMTs scenario)

- **Hit-only option**
  - 5,600 MB/s / 2000 FEM = 2.8 MB/s per FEM i.e. 25 Mbit/s per link (only?!!)
  - Per 48-port data concentrator: 48 * 25 = 625 Mbit/s. Fits in one 1 Gb Ethernet link

- **Waveform option**
  - 23,400 MB/s / 2000 FEM = 11.7 MB/s per FEM i.e. 100 Mbit/s per FEM link
  - Per 48-port data concentrator: 48 * 100 = 4800 Mbit/s. Fits in one 10 G Ethernet or PCIe Gen2x2

Figures and interpretation of the above table to be confirmed!

A POSSIBLE SYSTEM ARCHITECTURE



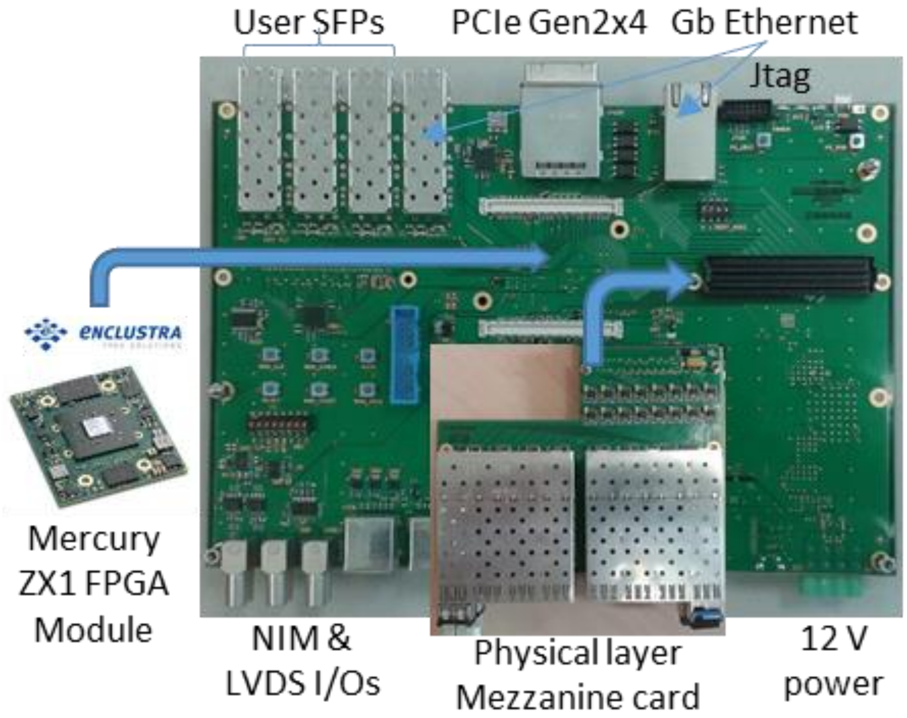
Primary Clock
GPS Module
Configuration Control Monitoring
Back-end Module
48 slave I/O ports
Optional fast links for main DAQ e.g. PCIExpress, Eth 1-10G
10-100 m optical fiber links
… up to 48
Standard Gigabit Ethernet Network
1 master I/O port
Back-end Module
48 slave I/O ports
… up to 48
1 master I/O port
Back-end Module
48 slave I/O ports
DAQ Servers & Network
125 MHz + 125 Mbps descend
500 Mbps ascend
~100 m optical fiber links
Distributed at surface level
In water tank

connect up to 48 x 48 = 2304 front-end modules, i.e. 55,296 PMTs assuming 24 PMT/module

Principles

- A dual stage fanout tree composed of custom back end-modules: 1 root and up to 48 leaves
- An ordinary Gigabit Ethernet network for global configuration, control and monitoring
- An optional fast data link from each leaf back-end module to a server in the main DAQ farm. Could offer redundancy for DAQ or merge clock + DAQ networks on the same front-end links

PRIOR DEVELOPMENT AT DEDIP: T2K HA-TPC BACK-END ELECTRONICS « TDCM »



D. Calvet, «Back-End Electronics Based on an Asymmetric Network for Low Background and Medium-Scale Physics Experiments », in IEEE Transactions on Nuclear Science, Vol. 66, N°7, pp. 998-1006, July 2019.

Trigger and Data Concentrator Module - TDCM

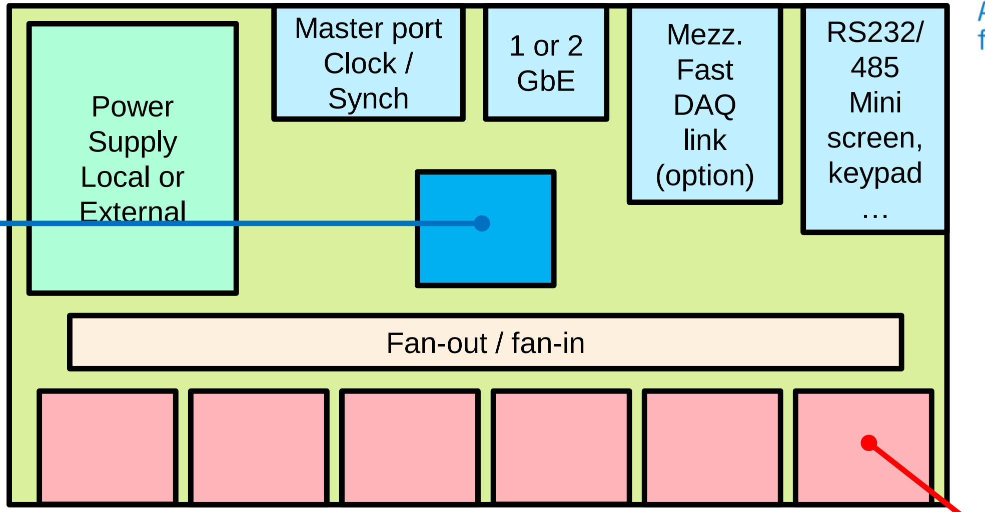
- Originally designed for PandaX-III; to be used for HA-TPCs readout in T2K upgrade and PUMA
- 1 Master port + up to 32 Slave ports. 1x100 Mbps from TDCM to front-end (reference clock, trigger, configuration) and 32 x 400 Mbps from front-ends to TDCM (detector data, monitoring)
- Low speed link to DAQ: Ethernet 1 Gbps RJ45 or GBIC (copper) / SFP (optical)
- High speed links to DAQ: 1-3 optical transceivers (6.6-10 Gbps) or PCI Express Gen 2 x 4 (16 Gbps net each way). High speed links currently untested – not used in T2K-II nor in PUMA
- Actual production cost: 3.8 k€ in 32 ports version equipped with 850 nm transceivers (10 TDCMs)

PROPOSAL FOR A DEMONSTRATOR OF BACK-END FOR HK CLOCK DISTRIBUTION

TE803 *249 €/u*

TE808 *914 €/u*

Xilinx Zynq UltraScale+, 2 GByte DDR4
(4 Gbyte DDR4 avec TE808)


Power Supply Local or External
Master port Clock / Synch
1 or 2 GbE
Mezz. Fast DAQ link (option)
RS232/ 485 Mini screen, keypad …
Fan-out / fan-in
48 slaves ports

Intel® Ethernet Server Adapter X520-DA1/X520-DA2 for Open Compute Project (OCP)

150 €/u

Stackable SFP 2x8

Main Features

- Based on Zynq Ultrascale+ SoC, Trenz TE803 (PCIe Gen2x4) or TE808 (PCIe Gen3x8)
- 1 Master port and 48 slave ports. 8x125 Mbps descend using CDCM (8x500 Mbps without)
  48 x 500 Mbps ascend – (1.25 Gbps per port reachable from datasheet, but let’s check first…)
- Low speed 1 GbE for control or slow DAQ
- Mezzanine (OCP standard?) optional for high speed DAQ: Dual or Quad Ethernet 1G, Ethernet 10G, PCIe Gen 2 x 4 on Samtec Firefly (up to Gen3 x 8 with TE808)
- Possible format: 6U x 14F (6 per 6U crate) or 1U x 84F (same form factor as WR3/4 switches)
- Estimated cost: ~4 k€ with clock distribution only, ~5k€ with 2 Ethernet 10 G ports
- Total back-end boards cost (for 48,000 PMTs) : 172-215 k€ (without spares and contingency)

DEMONSTRATOR GOALS

Demonstrator Goals

- Demonstrate if 48-port density is reached on single board and what link speed is achieved
- Perform R&D on CDCM technique to evaluate its performance and applicability
- Investigate an asymmetric topology / bandwidth network
- Evaluate clock distribution performance in a realistic setup scalable to the full required size
- Build a distributable base platform for firmware and software development while the final hardware version is being designed (assuming that scheme is chosen)
- Gain experience with high speed interfaces (≥10 Gbps) for optional DAQ support
- Alternative design to proven or not yet proven options
- Not aimed to be a final board

INVOLVEMENT STRATEGY

My view on the strategy for DEDIP

- Late joiners in a project that has already been structuring itself for >18 months
- No benefit for anyone that we duplicate work already done, e.g. evaluation of White Rabbit
- No intention to compete on a custom solution based on traditional FPGA SerDes scheme
- Propose to engage in more disruptive – and risky – ideas and implementations
- Open to collaborate with interested parties to the development of this option
- See positively (less risks) that different designs are explored in parallel until one is selected
- Shall ensure, or ask that it is ensured, that each participant group gets a sufficient share of the work once final choices have been made – although selecting one solution will mechanically bring its proponents on the forefront
- Shall commit to provide a fraction of the required funding independently of the solution that will be retained