Status Updates on the PXD-DAQ -
Setup for Upcoming Test Beams and Monitoring

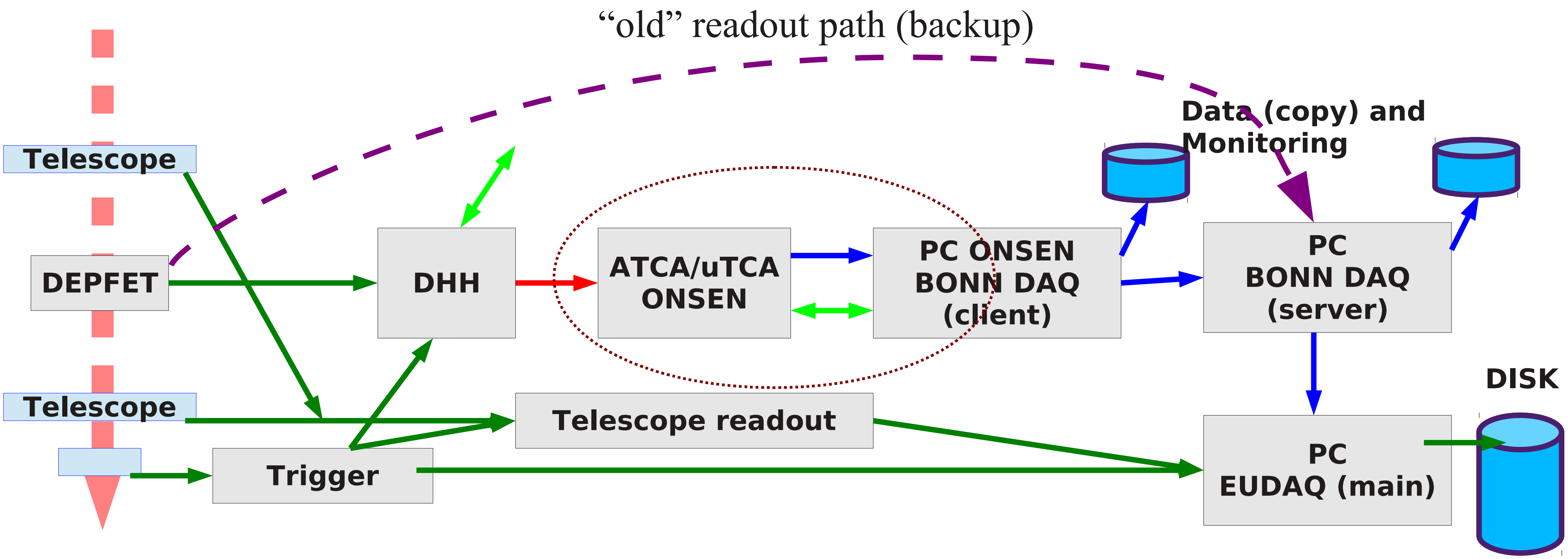
Björn Spruck
(for the Giessen Group)

JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

Belle II

DESY Test May 2013

Concept for May 2013 DESY Test

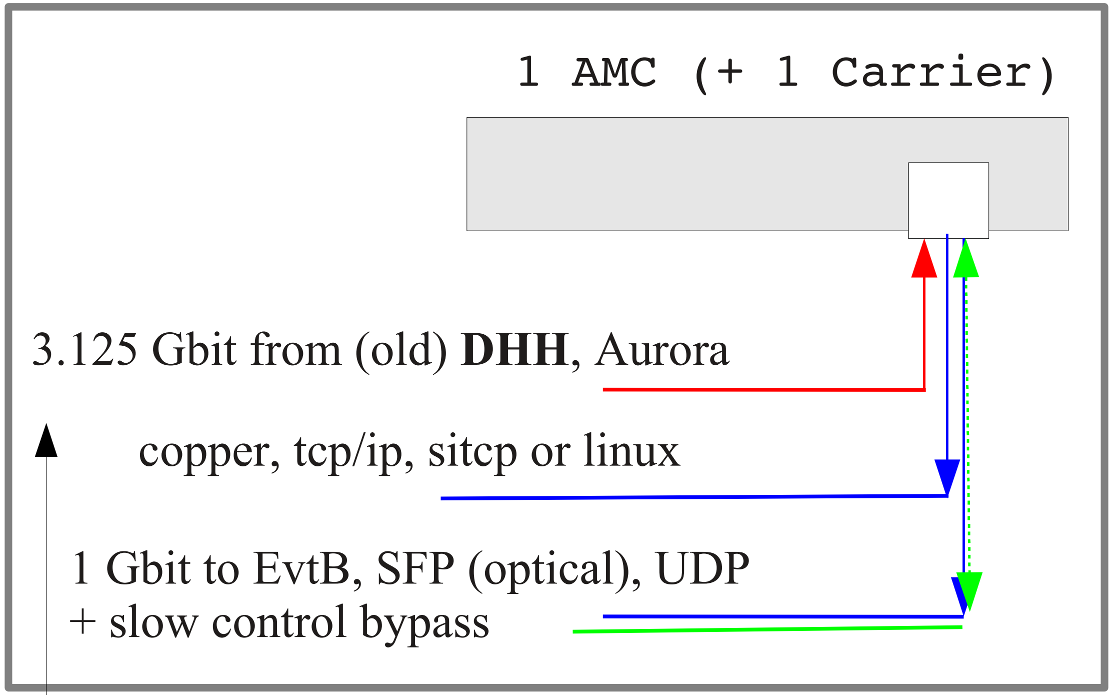


- Connect ONSEN to EUDAQ by "Bonn DAQ"client
  - Bonn DAQ is present anyway (backup)
  - Monitoring for DEPFET is done in Bonn DAQ
  - Bonn DAQ can be tested standalone (without Trigger)
- DHH has to add trigger number to data stream

Concept for May 2013 DESY Test

- For ONSEN: “Just” data readout
  - No HLT trigger, no ROIs
  - Data In -> (buffering) -> Data Out
  - Minimal slow control
  - No IPMI (needed)
- Stripped down firmware
  - No ROI selection logic in FPGA
- Aurora receiver, buffer management logic, SITCP and/or UDP sender, Linux based tcp/ip is not needed.
- UDP more challenging on receiver side (BonnDAQ producer), TCP more strait forward.
- Hardware wise no problem to keep both cores.

As carrier board is passive, no SC is needed.
We can use pizza-shelf or even a uTCA shelf without carrier board (tbd).



1 AMC (+ 1 Carrier)

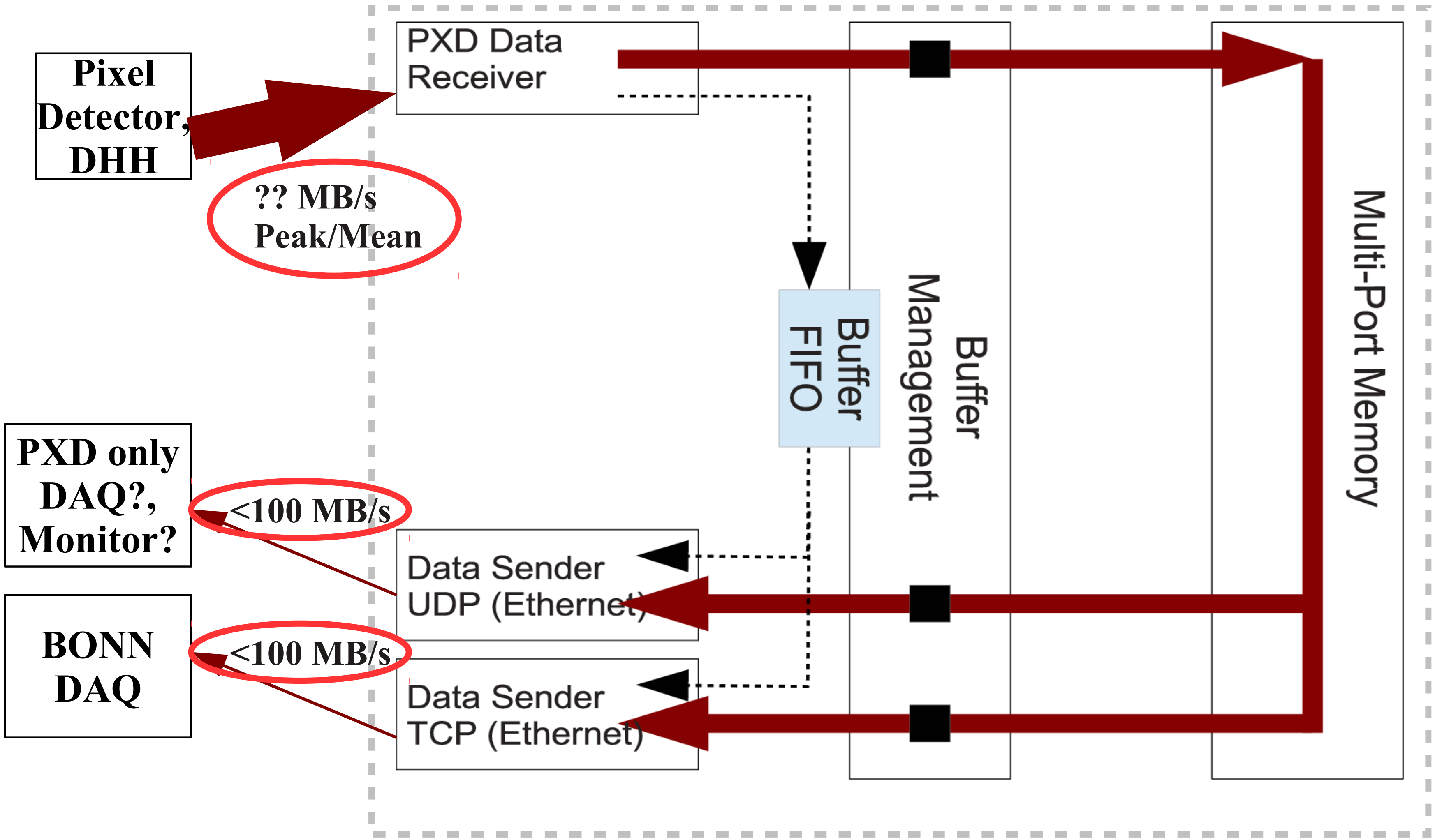
3.125 Gbit from (old) **DHH**, Aurora

copper, tcp/ip, sitcp or linux

1 Gbit to EvtB, SFP (optical), UDP
+ slow control bypass

3.125, decided today to allow
DHH to switch to PCIe card for debugging

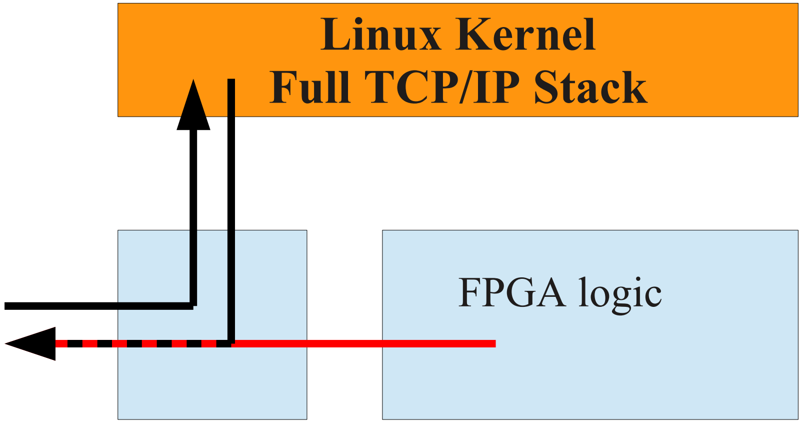
On Chip Data Flow for the PXD Data, DESY Test 2013



- Data in format: DHP/raw, zero/not-zero suppressed
    - Chained buffers, as event data size is variable (backup: fixed size)
- Data out is duplicated, keep TCP as backup, test UDP

UDP Implementation on ONSEN

- UDP core successfully ported to Virtex 5
  - SFP and on-board Gbit Ethernet tested
- New: TCP bypass/tunnel to Linux kernel
  - Can do slow control on same physical line
  - Until now, no dedicated kernel driver, done by TAP network interface
  - ARP, DHCP handling can be handled either in HW or SW (switchable)
- But: Pause Frame not (yet) working
- TODO: Verify SFP RJ45 (copper) is working as well! **It does!**
  - ~~Check two (cheap) transceivers, they require SGMI interface, but up to now we used 1000X and GMII only.~~
  - ~~Small modifications on IP core are needed~~

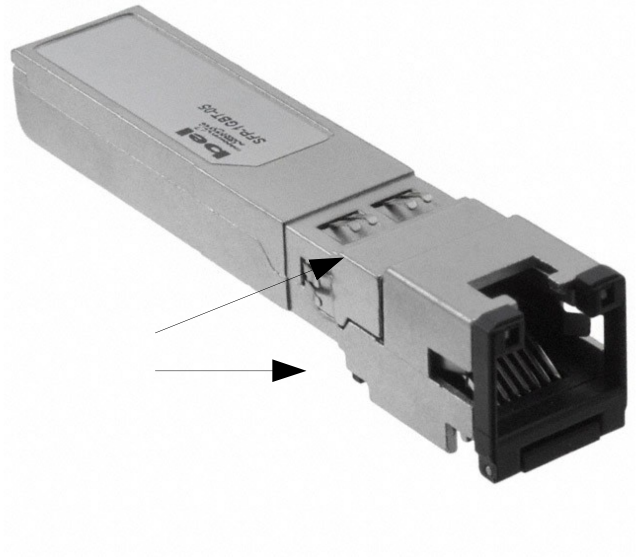


UDP Out from HW

It turned out, the FINISAR Transceiver worked out of the box. Without any core modification!

Additional Gbit Ethernet on SFP transceivers

- Optical SFPs (Finisar) were working on Virtex 5
- Bought three “copper” GbE (RJ45)
  - Bel Fuse SFP-1GBT-05 → protocol issue, needs more effort
  - AVAGO ABCU-5700RZ → protocol issue, needs more effort
  - Finisar FCLF-8521-3 → worked out-of-the-box (with 1000X protocol same as optical transcever; 60,- Euros)
- The Finisar Transceiver is not overlapping the other channels of the cage!



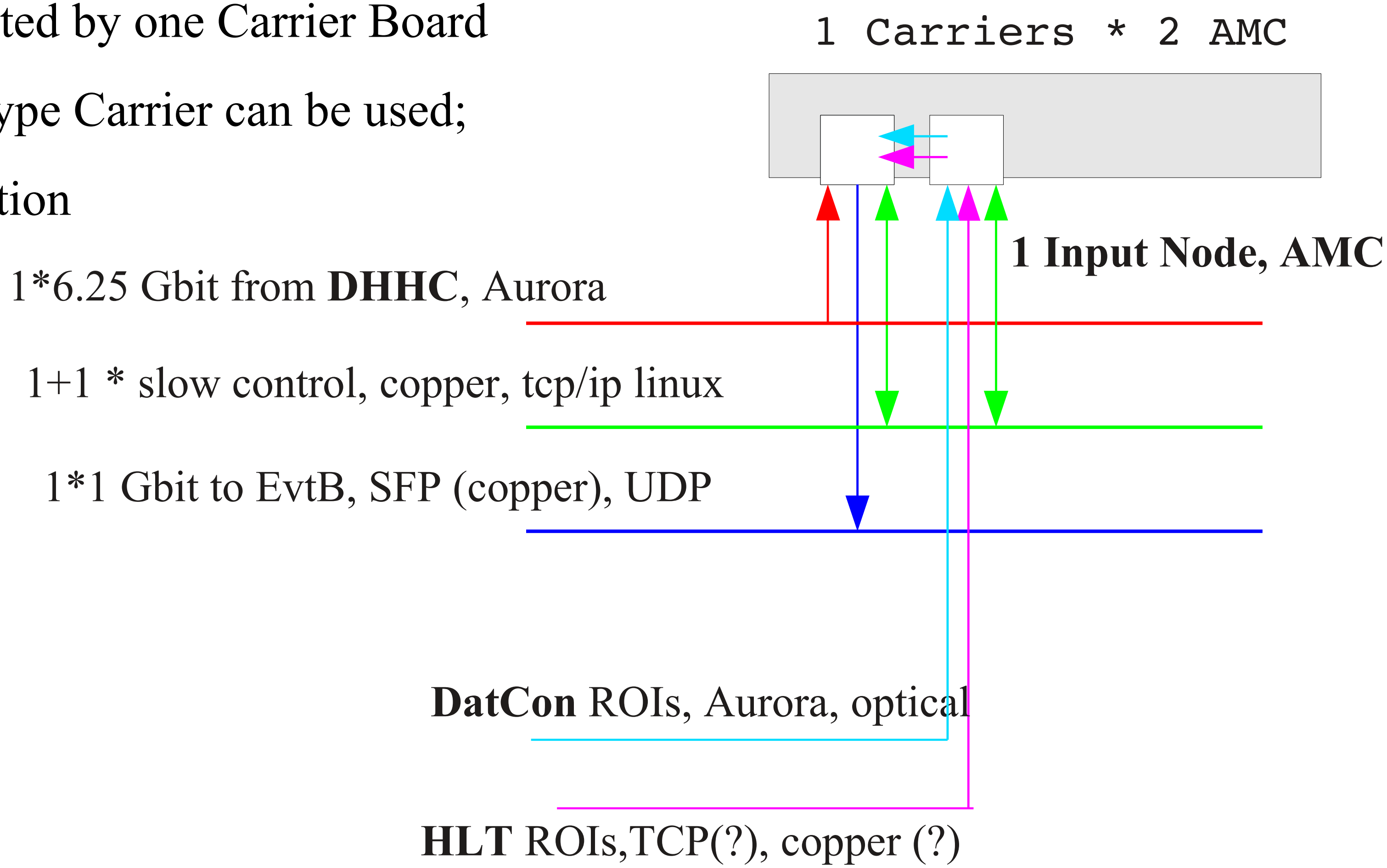
Monitoring (also DAQ)

- Basic ONSEN:
  - No IPMI. No monitoring of board health (beside maybe FPGA temperature)
  - No EPICs. There is not much to monitor, neither to change states.
  - Buffer usage, Data rate, Frame rate, Aurora errors, Network (TCP sender), …
- Basic PXD on ONSEN
  - Would require depacking of data, not foreseen for this test
- PXD monitoring in BonnDAQ … exists, has to be adopted
- Request from DHH, have a RAW data (memory dump) after pedestal update <u>outside</u> of BonnDAQ
  - For check if pedestal upload worked flawless (glitch in writing registers)
  - Immediately, data delivert to DHH "pedestal upload software"
  - Split this data path either at the BonnDAQ producer <u>or</u> use separate data path (which is foreseen for UDP transmission anyway)

DESY Test Jan 2014

Concept (For 2013/14 DESY Test)

- Aim: Close to final readout chain
  - PocketDAQ, EPICs Slow Control
  - High Level Trigger, DATCON → ROI handling
- Only on DHH → 1 AMC for ROI selection
- 1 AMC Input Node, connected by one Carrier Board
- No backplane usage, prototype Carrier can be used;
  - uTCA as backup solution



1 Carriers * 2 AMC

1 Input Node, AMC

1*6.25 Gbit from **DHHC**, Aurora

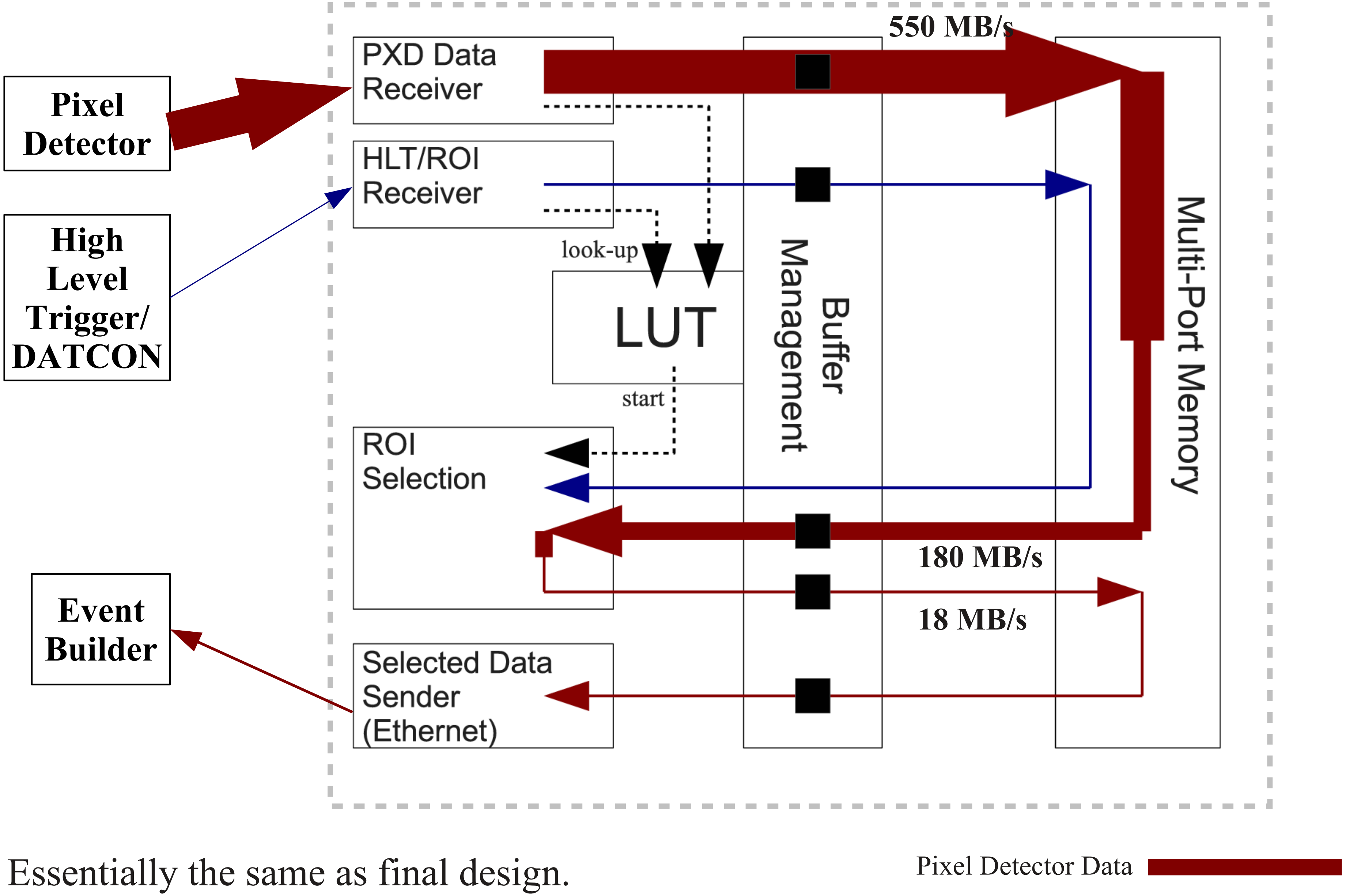
1+1 * slow control, copper, tcp/ip linux

1*1 Gbit to EvtB, SFP (copper), UDP

DatCon ROIs, Aurora, optical

HLT ROIs,TCP(?), copper (?)

On Chip Data Flow for the PXD Data (DESY 2014)



Essentially the same as final design.
Less complex in terms of (sub-)event building
Differences in Input and Output

Differences to Final System

- 2 halfladders (with DHHC: 5)
  - Reduces lookup complexity
- Output: no sub event building
- No (active) carrier board connection.

Monitoring (also DAQ)

- Basic ONSEN:
  - IPMI. Monitoring of board health, power etc
  - EPICs. For Buffer usage, Data rate, Frame rate, Aurora errors, Network etc
- Basic PXD on ONSEN
  - Anything beside “simple” counting is hard to do:
  - Example HitMap require random access to system memory in addition to all other components (additional RAM interface), because of RAM access time this cannot be directly added to the ROI core. Second data depacking path would be needed... and so on...
- Request from DHH, have a RAW data (memory dump) after pedestal update … how are requirements for Desy 2014? Most likely not.
- PXD monitoring should be done on PC (next slide)

Monitoring (also DAQ)

- One point: We do not want to avoid implementing (complicated, time consuming) monitoring which is only used in DESY 2014
- Implement it such, that it can be reused!
- From VXD DAQ/DQM/SC Software EVO Meeting 28 March 2013:
  - “With contributions from Alan, Björn, Christian, Peter and Sören, possible implementation scenarios have been discussed. The preferred way to implement our monitoring is to run it on an intermediate PC that also does the UDP-to-TCP conversion. Implementing it within the basf2 framework right from the start would be interesting, but requires more manpower.”
  - “The ONSEN system will also provide monitoring information (data reduction factors, FIFO fill factors, etc.). “

Data Format ONSEN → DAQ

- Problem: DHH format does not include the length of the packets, neither does DHP.
- There are some frame types with variable length
- Therefore: Frames need additional encapsulation
- ONSEN knows the size of the frames, as this is recorded when they are written to memory.
- Minimal: Magic (32 bit), Length (32bit)

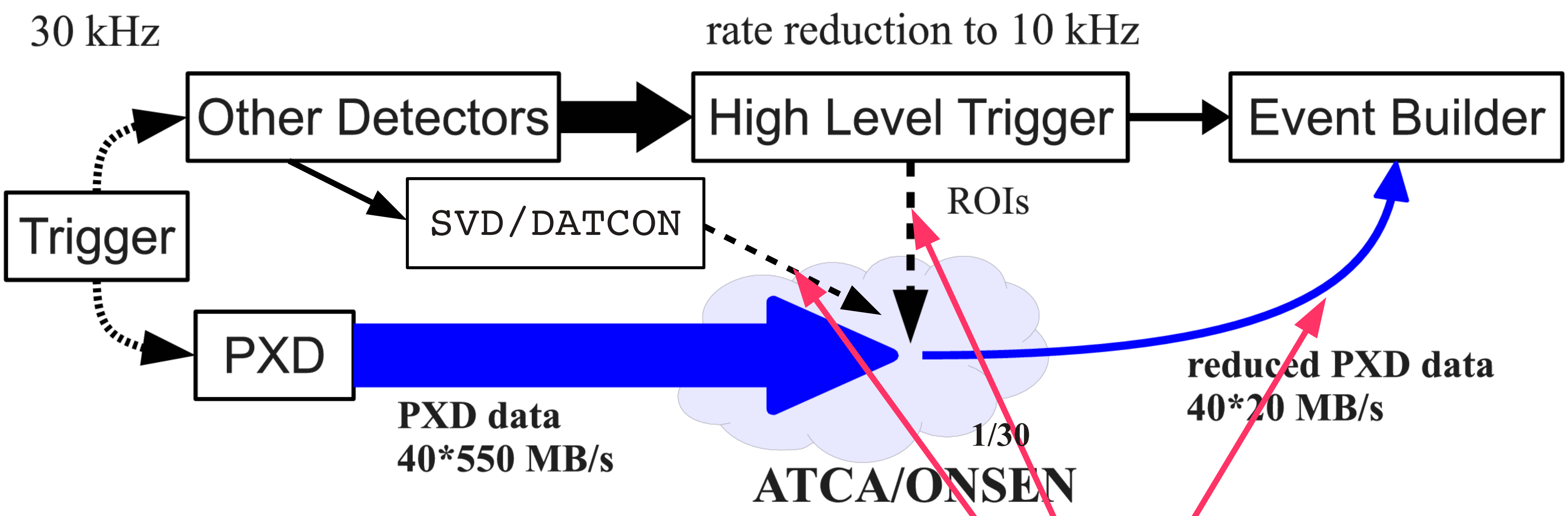
Backup

Connectivity Summary: DAQ, RC, SC, ...

- Front panel:
  - 32 Optical lines DATA IN from DHHC, Aurora 1:1
  - 32 Ethernet DATA OUT to UDP2TCP box, UDP 1:1
  - 32 Ethernet Slow Control/Run Control/Monitoring (TCP) from switch
- By backplane:
  - Stream of ROI/Trigger info from Input Node, Aurora
- Input Node:
  - 1 Ethernet HLT DATA IN (UDP or TCP?)
  - 1 Optical DATCON DATA IN to input node, optical 1 Gbit or more
  - Both switched to RocketIO on full mesh backplane, Aurora
- Ethernet Switch:
  - 1 link Slow Control/Run Control for Carrier Boards

(for more detailed discussion, see
my talk from Seoul 2013 DAQ workshop)

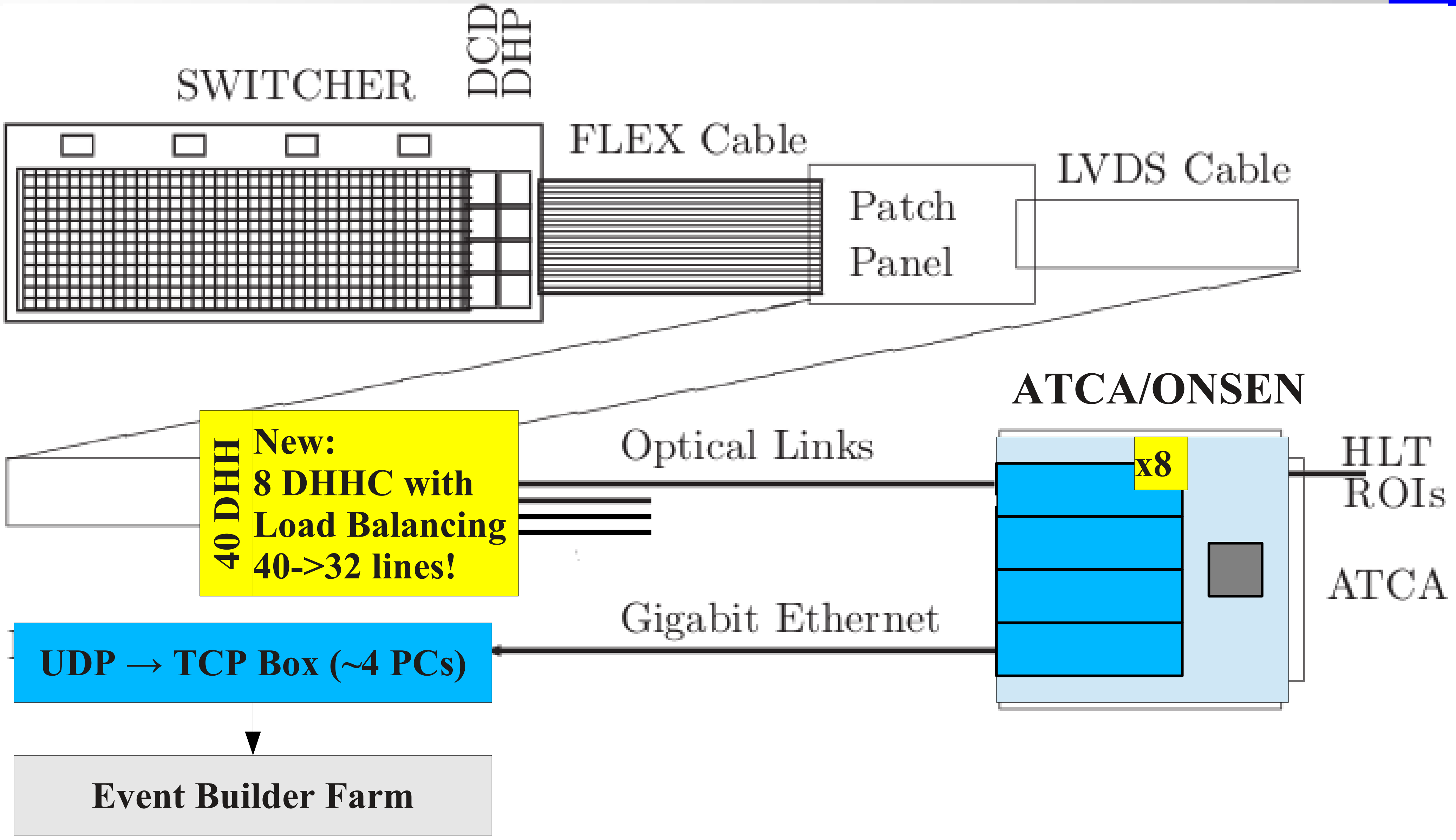
Reminder: Pixel Detector DAQ Scheme



Physical layer and protocol
TCP vs UDP, Aurora
optical vs copper

- TCP/IP wanted by Event Builder
- UDP preferred from FPGA side
- UDP unsave, packet loss on switches?
- Optical or copper, cable type and how much!
- Put a UDP ->TCP converter box in EHut
    - (Feasibility study: See talk of Yamagata-san on Seoul DAQ workshop Jan 2013)

Reminder: PXD Readout Chain (Updated)

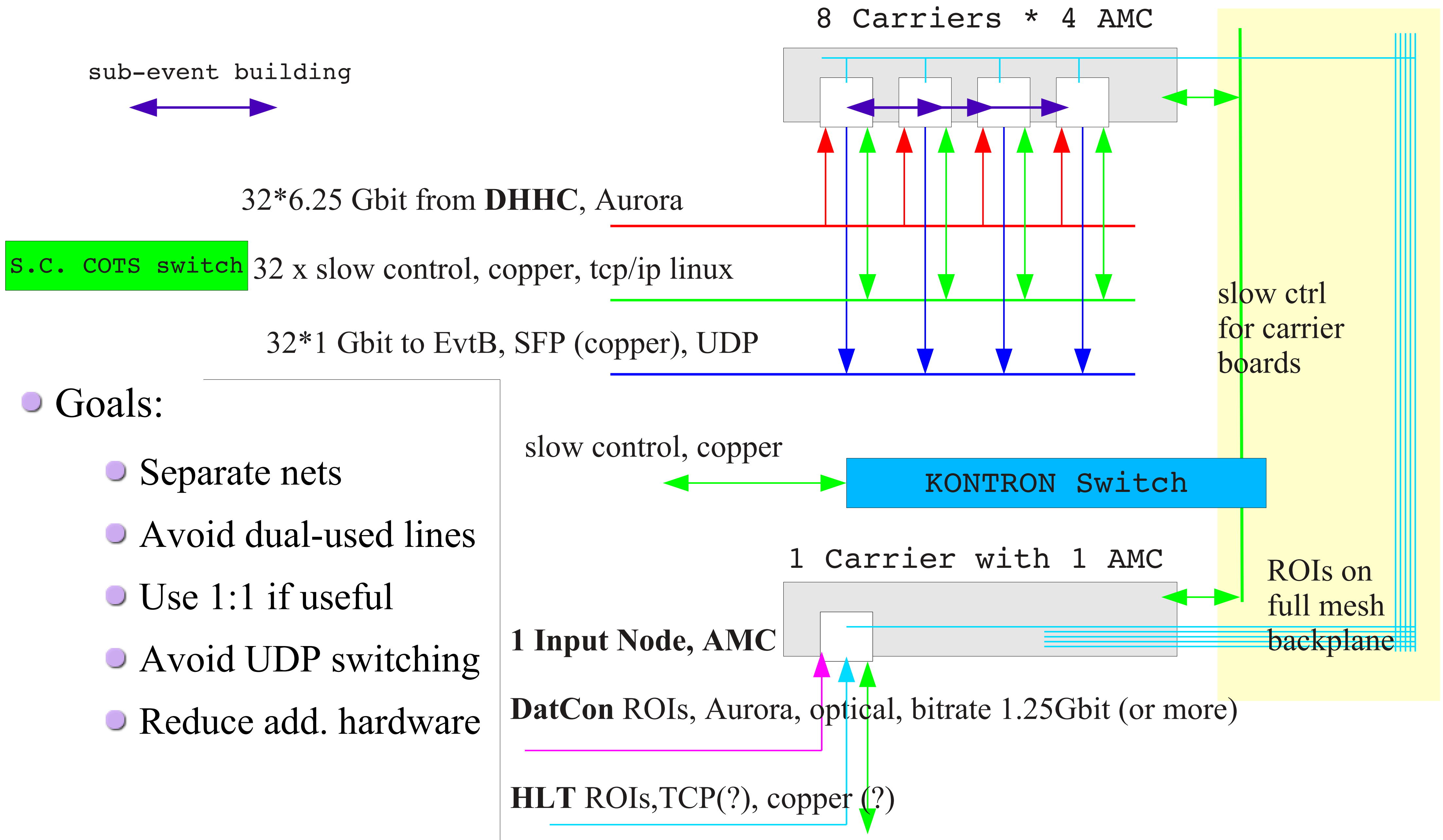


New:
8 DHHC with
Load Balancing
40->32 lines!

UDP → TCP Box (~4 PCs)

Event Builder Farm

Concept (Discussed and Approved in Seoul)



8 Carriers * 4 AMC

sub-event building

32*6.25 Gbit from **DHHC**, Aurora

S.C. COTS switch 32 x slow control, copper, tcp/ip linux

32*1 Gbit to EvtB, SFP (copper), UDP

slow ctrl
for carrier
boards

slow control, copper

KONTRON Switch

1 Carrier with 1 AMC

ROIs on
full mesh
backplane

1 Input Node, AMC

DatCon ROIs, Aurora, optical, bitrate 1.25Gbit (or more)

HLT ROIs,TCP(?), copper (?)

- Goals:
  - Separate nets
  - Avoid dual-used lines
  - Use 1:1 if useful
  - Avoid UDP switching
  - Reduce add. hardware

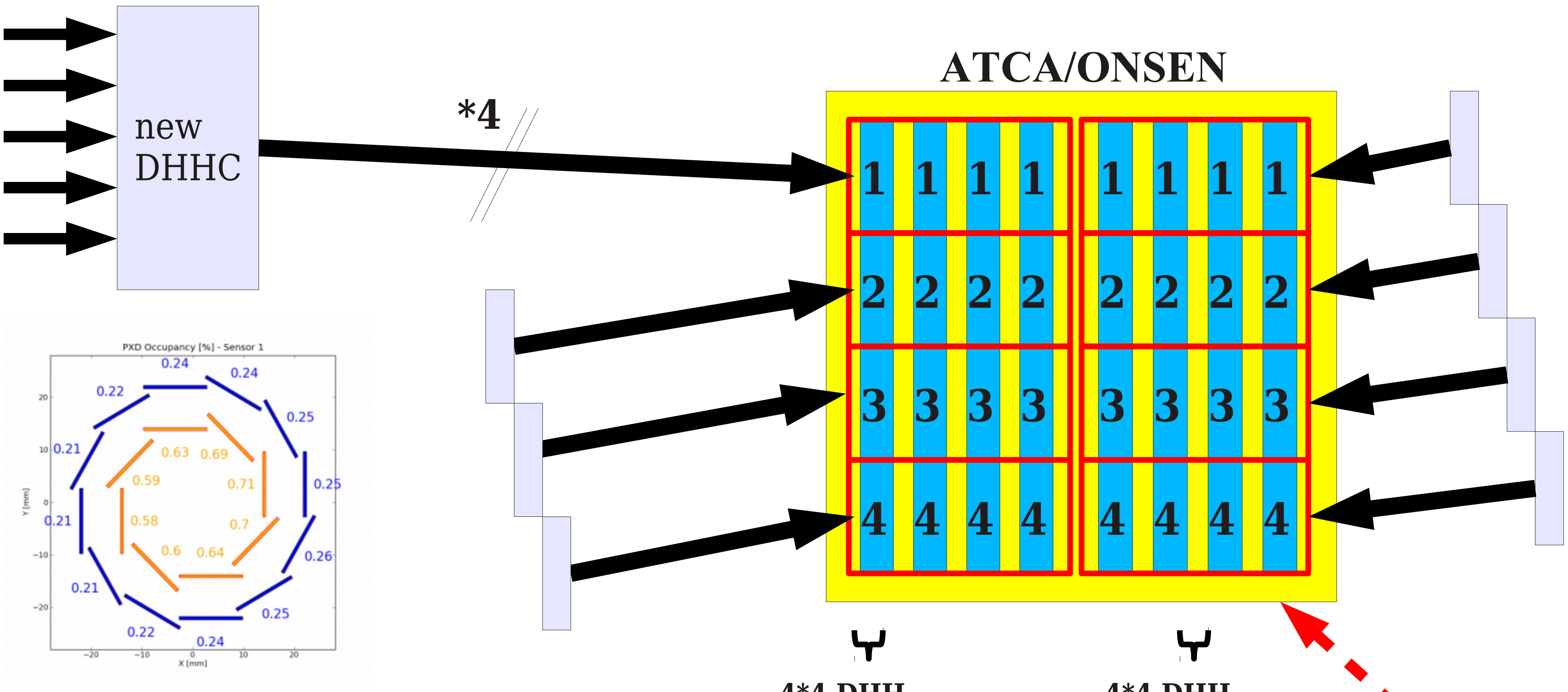
ONSEN (Test Beam) Setups

- Towards ONSEN final setup
  - Cabling issues
- Setup for DESY Jan 2014 (was planned for Oct 2013)
  - incl DATCON, HLT, Pocket DAQ
- Setup for DESY May 2013
  - only PXD, EUDAQ

Summary and Outlook

- Trying for a final cable and protocol setup
  - Still open questions (mainly UDP)
- Test showed UDP is feasible, additional PCs needed
  - Locate PCs in E-Hut next to ONSEN
    - Saves (optical) cables from E-Hut to Computing Room
- Proposal for DESY 2014 setup
- Setup for DESY test in May 2013

New DHH/DHHC -> sub- and full event building?



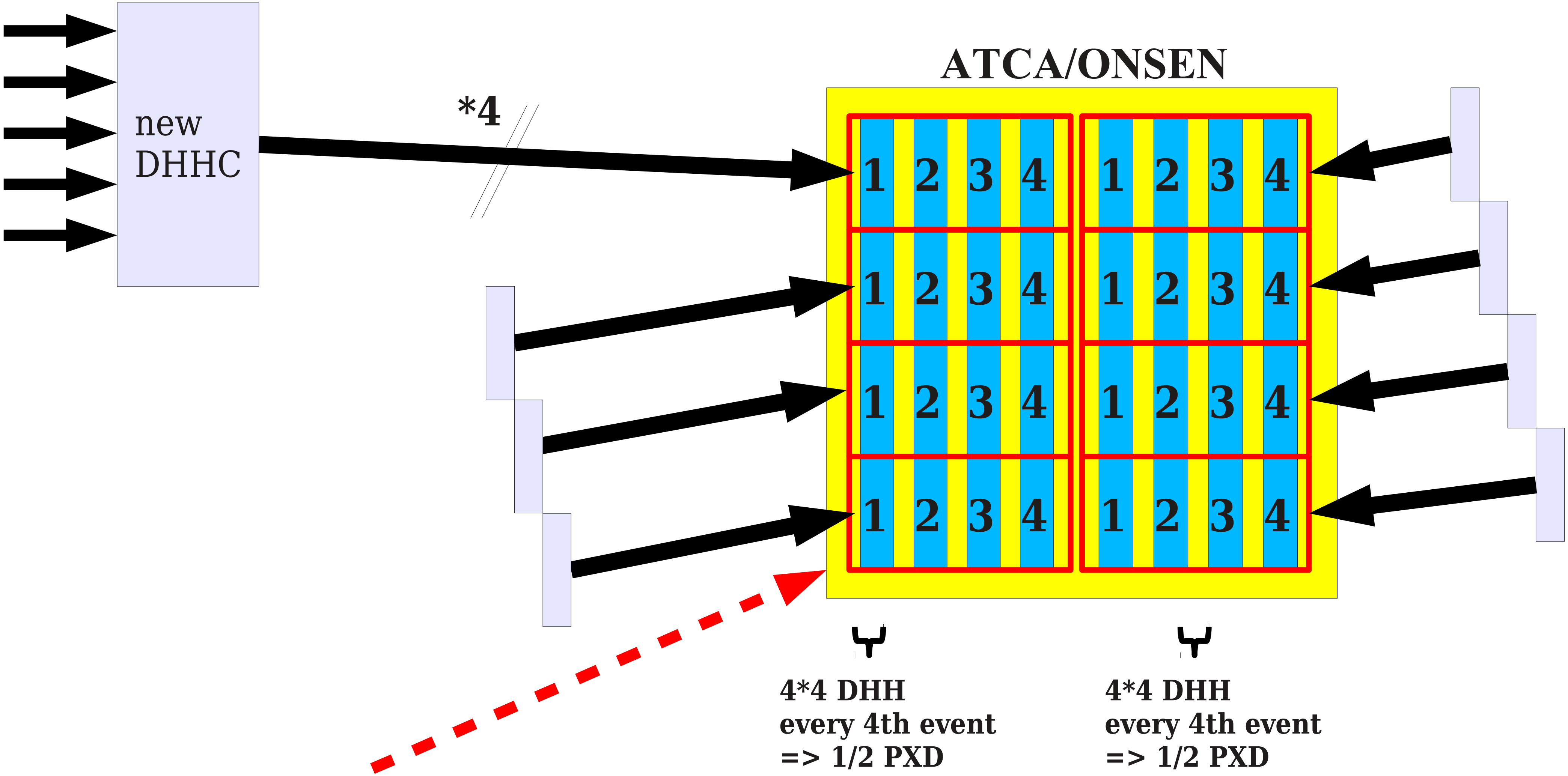
4*4 DHH
every 4th event
=> 1/2 PXD
f.e. upstream part

4*4 DHH
every 4th event
=> 1/2 PXD
f.e. downstream part

40 Half modules
(2+3) load balance
8 DHHC modules -> each 1/8 of PXD
1 DHHC -> 4(!) ONSEN daughter boards
=> **option**: subevent building on mother board
for 1/2 PXD event

These are not Event Numbers!

Event Wise View



4*4 DHH
every 4th event
=> 1/2 PXD
f.e. upstream part

4*4 DHH
every 4th event
=> 1/2 PXD
f.e. downstream part

Now, these <u>are</u>
Event Numbers!

Connectivity to DAQ, RC, SC, ...

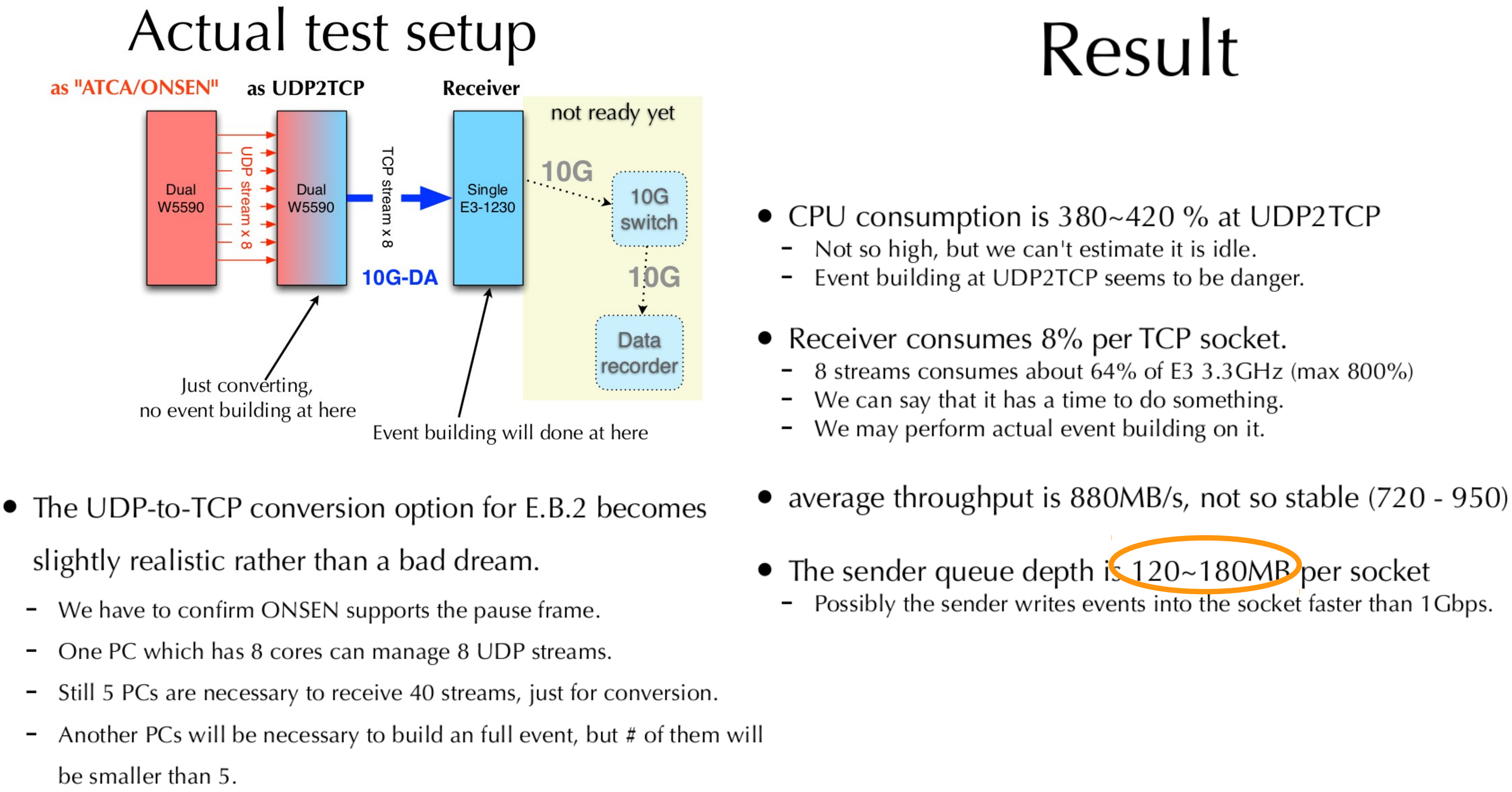
How many different networks do we have to serve?

- 32 DATA out – direct connections (UDP)
- HLT data in (UDP or TCP?)
- (DATCON data in, most likely optical on an additional AMC board)
- Slow Control/Run Control/Monitoring (TCP)

- Until now: HLT, DatCon and SC input not precisely defined
  - Argument: Rate low -> can be done on backplane ethernet
  - Forseen for DatCon: Input Node -> switch to Rocket-IO on Backplane
    - From there, forward to AMCs

UDP -> TCP Box

- Yamagata-san presented result on UDP->TCP performance



Actual test setup

Just converting, no event building at here

Event building will done at here

- The UDP-to-TCP conversion option for E.B.2 becomes slightly realistic rather than a bad dream.
  - We have to confirm ONSEN supports the pause frame.
  - One PC which has 8 cores can manage 8 UDP streams.
  - Still 5 PCs are necessary to receive 40 streams, just for conversion.
  - Another PCs will be necessary to build an full event, but # of them will be smaller than 5.

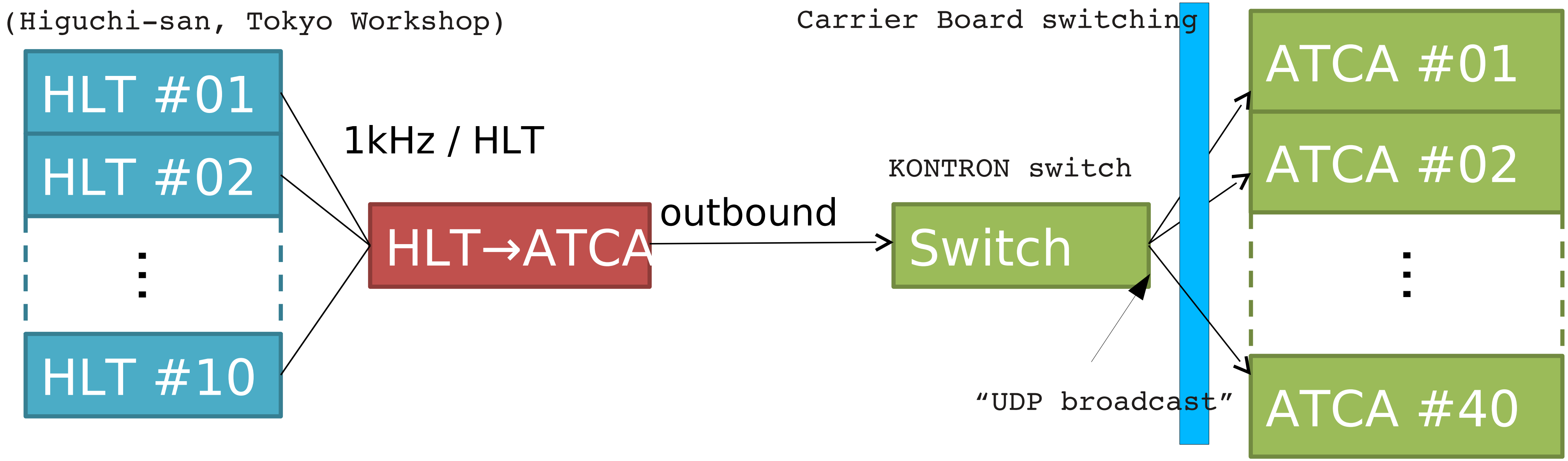
Result

- CPU consumption is 380~420 % at UDP2TCP
  - Not so high, but we can't estimate it is idle.
  - Event building at UDP2TCP seems to be danger.
- Receiver consumes 8% per TCP socket.
  - 8 streams consumes about 64% of E3 3.3GHz (max 800%)
  - We can say that it has a time to do something.
  - We may perform actual event building on it.
- average throughput is 880MB/s, not so stable (720 - 950)
- The sender queue depth is 120~180MB per socket
  - Possibly the sender writes events into the socket faster than 1Gbps.

Remarks

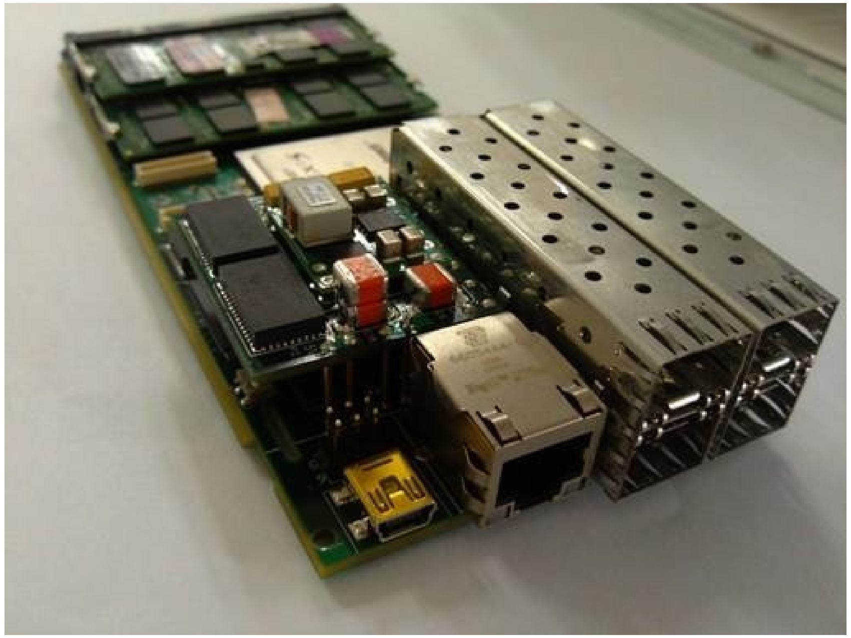
- 4 PCs
  - UDP2TCP as well as 1 GBit -> 10 GBit
  - Reduce # cables 32 -> 4
- By putting the UDP2TCP PCs in E-hut, number of outgoing lines is reduced
  - Enough rack space? Yes!
- Yamagata-san is estimating worst case
  - Our estimations were 20-30 MB/s/cable (550/30+unsuppressed data)
  - We have already sub-events
    - 5 ladders together because of DHHC load-balancing
    - 5*4 option on one carrier board w/o using backplane by clever cabling
      - half of the detector (up or downstream)

HLT Connection (First Idea ...)



- one cable for 30kHz? => 30k UDP packets??? means 10kHz ROIs + 20kHz*20 bytes payload + IP frame/header
- tough! 62MB/s without overhead (and without DatCon!)
- HLTs are "broadcast on the backplane" => very inefficient
- Motherboards have to process 62MB and distribute it to the correct AMC
  - Requires hardware programming for the carrier board UDP(?) packet processing and switching (not started yet)
- Do it more intelligent? Do real switching?

AMC Rev 2



- So far two revisions. Both had:
  - Virtex-5 FX70T FPGA. (Could be exchanged for smaller, pin-compatible one.)
  - 4 GB DDR2, Gigabit Ethernet, 64 MB Flash, UART to USB converter, . . .
- Rev 1->2: Bugfixes
- AMC connector pinout was revised.
- The second revision has 4 instead of 2 SFP ports, as requested for DATCON.
- Rev 2->Rev 3: minor changes, now in production
- But: Carrier board needs redesign.

- Consequences for ONSEN:
  - Same board design as the 2 cage version but additional connectors!
  - F.e. additional ethernet if needed (with optical or copper SFP connector).
  - Each two cages share the same clock -> f.e. 2*6.25Gbit and 2*1.25Gbit

Remarks, Cables & Places

- ONSEN Input: 32 optical
  - + HLT, DATCON, 40*SC
- ONSEN Output: 32 ~~optical~~ copper
- DHH located @ detector
- ONSEN located @ E-hut (too long and too many joints for optical DHH -> Computing Room)