Belle II HLT Status and Plan

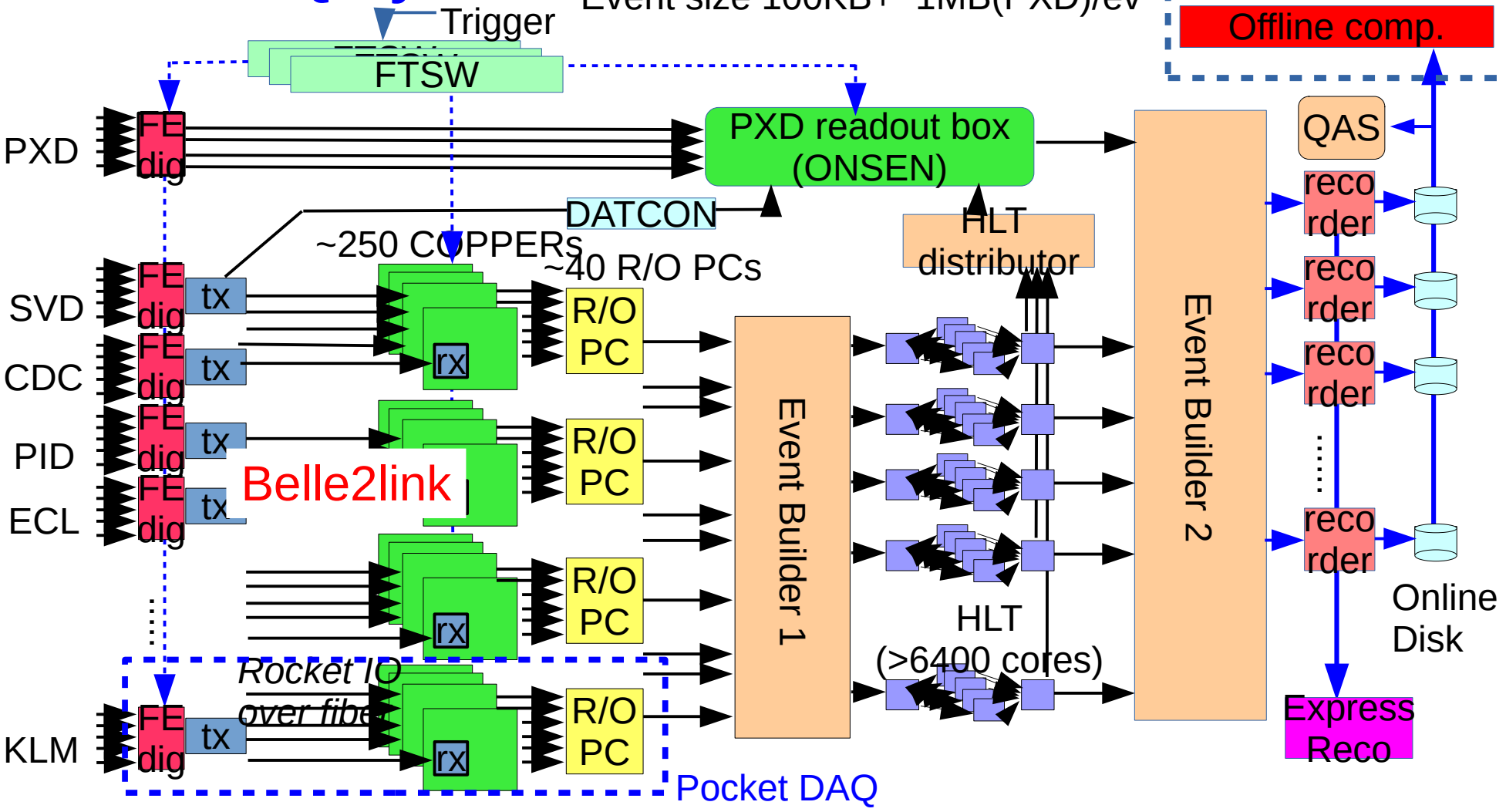
and

some thoughts on slow pion tracking in Belle II DAQ

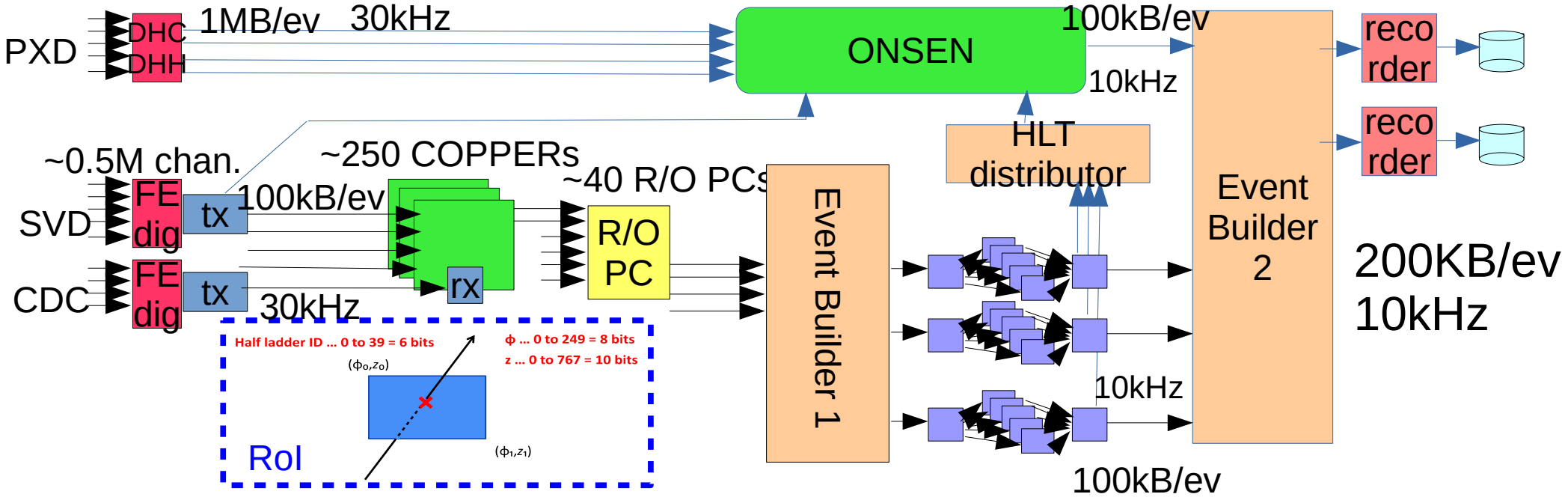
R.Itoh, KEK

Belle II DAQ System

Maximum design rate = 30kHz
Event size 100KB+~1MB(PXD)/ev



Readout for Pixel Detector (PXD)



- PXD yields a large event sized data when occupancy is high (>1MB) but it cannot be processed by COPPERs, nor recorded without event reduction.
- Data size reduction by 1) extrapolate HLT-reconstructed tracks to the surface of PXD sensors (Region of Interest), 2) send the RoIs to PXD readout box, and 3) discard hits not in RoIs. -> 1/10 reduction is expected.
- RoIs are sent only for HLT-selected events, and the rate reduction is also applied.

1. Hardware of Belle II HLT

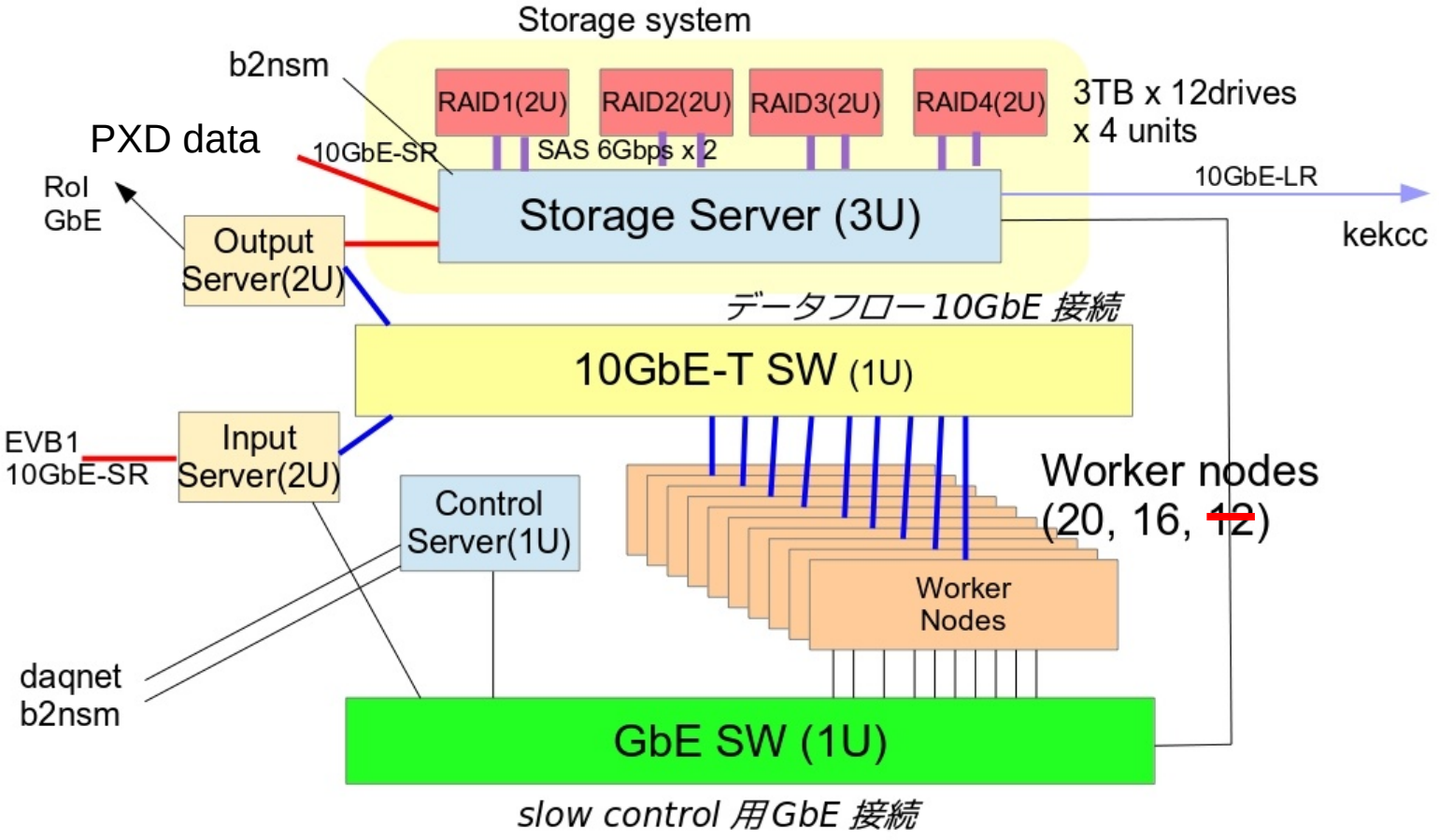
- Unit structure: Coarse parallelism is implemented by the unit structure. Event builder distributes the events to each unit following the modulo of event number.

- One unit consists of an input server, an output server, and up to 20 worker nodes with a control server.

- Each worker is equipped with muticore CPU(s) providing 16-40 physical cores / server.

- Data flow nodes are connected via 10GbE network while all nodes are via GbE network for the system control.

Actual Implementation of HLT/Storage unit



Equipped in a single rack

History and Plan of HLT reinforcement

just scaled by number of cores.

	No. of HLT units	No. of cores	Max. Rate	Remarks
Until 2020 summer	9	2880	8kHz	
2020-2021	10	4800	>12kHz	* More servers in each unit * OS upgrade to CentOS7
2021-2025	11->15	4800->6400	12-20kHz	* Yearly addition of units
2025-	15	>6400	>20kHz	For full Lum.

2020 HLT Reinfocement (summer+winter)

HLT01:
  16 cores * 9 + 20 * (2+2) + 28 * 2 + 36 * 2 + 40 * 3 = 472 cores
    (replace 6 servers with new ones).
HLT02-05
  20 cores * 16 + 36 cores * 2 + 40 cores * 2 = 472 cores
HLT06-09
  28 cores * 12 + 36 cores * 2 + 40 cores * 2 = 488 cores
HLT10 (new)
  28 cores * 12 + 36 cores * 2 + 40 cores * 2 = 488 cores

2880 cores → 4800 cores

~2000 cores have been newly added.

- The reinforcement achieved 75% of the design number of cores(6400).
- At the same time, the operating system has been upgraded to CentOS7.

Reinfocement schedule in coming years

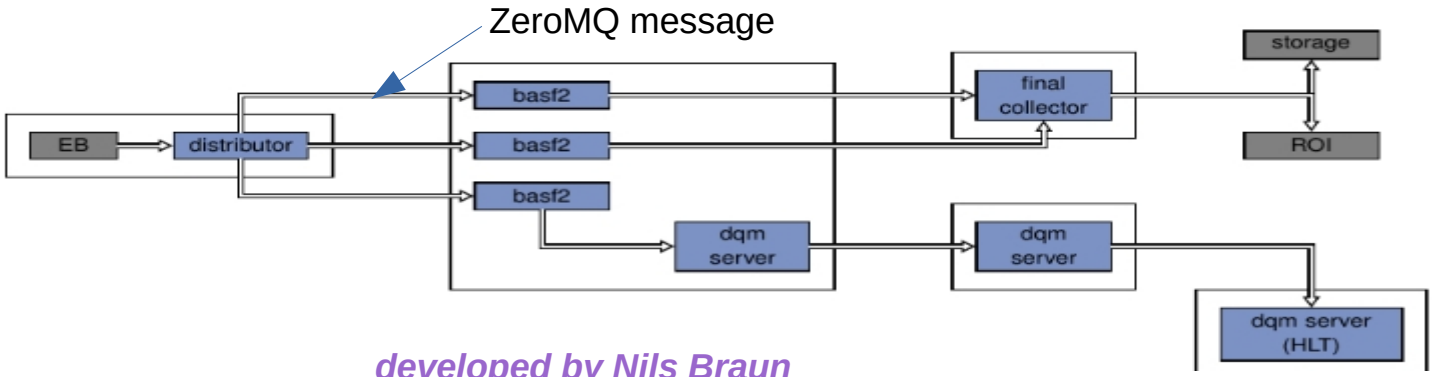
HLT:

- 3 units with 480*3=1440 (or more) cores will be added during next summer shutdown.
  -> The total number of cores becomes close to the design number (4800+>1440 <->6400).
- Further reinforcement plan should be discussed by looking at the increase in trigger rate/luminosity in 2021c/2022a run.
  => Plan should be decided before LS1.
- In addition, the replacement of old servers should be started. The 1st HLT unit (HLT01) was built in 2014 and the servers are already obsolete....
  => The reinforcement plan should include the replacement.

2. HLT software framework

Framework was switched to new ZeroMQ based system

OVERVIEW (A BIT SIMPLIFIED)



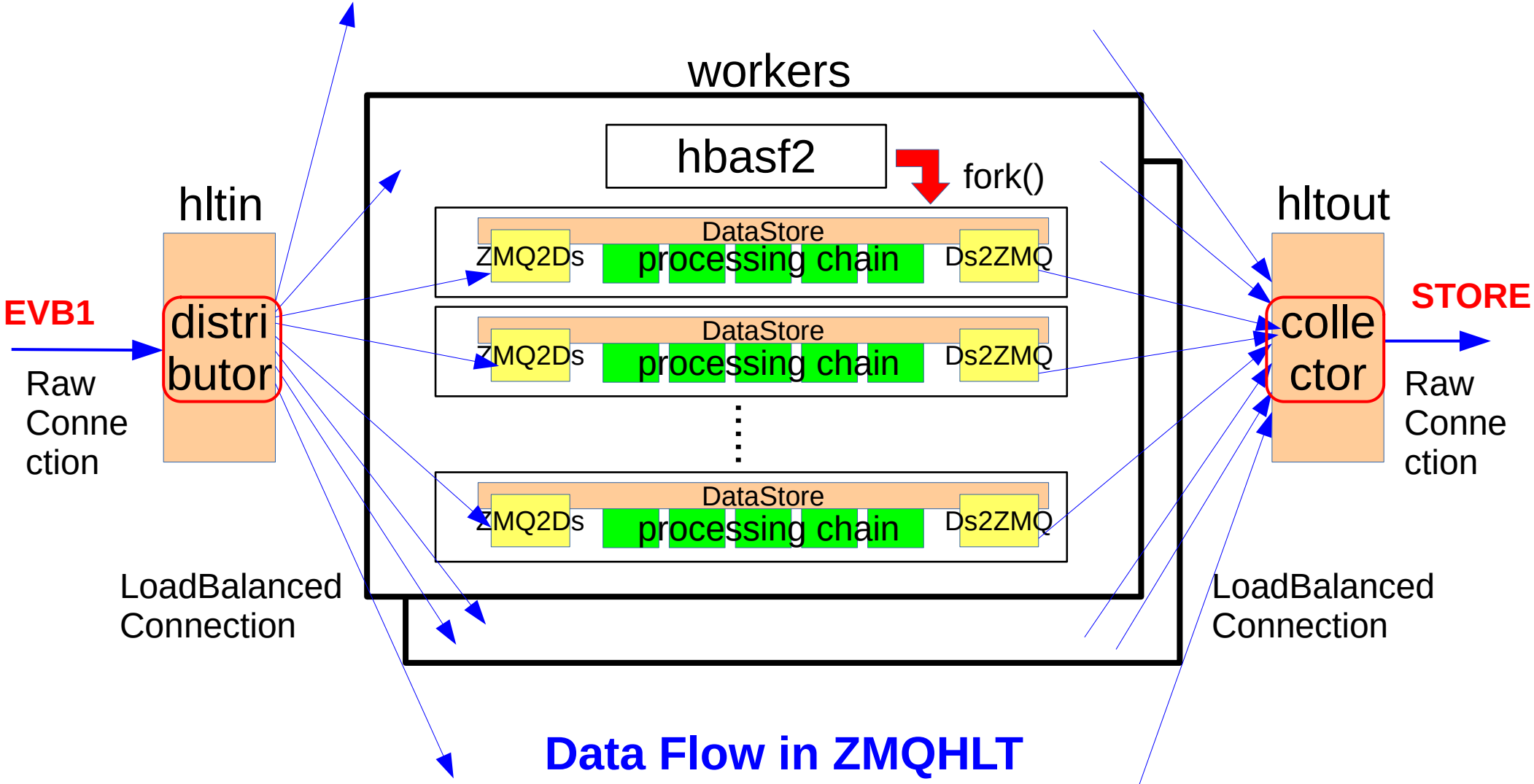
developed by Nils Braun

- RingBuffers are replaced with ZeroMQ message transport.

- Initialization of processing is done when making the system ready (not at the time of receipt of the first event) by using modified version of basf2.

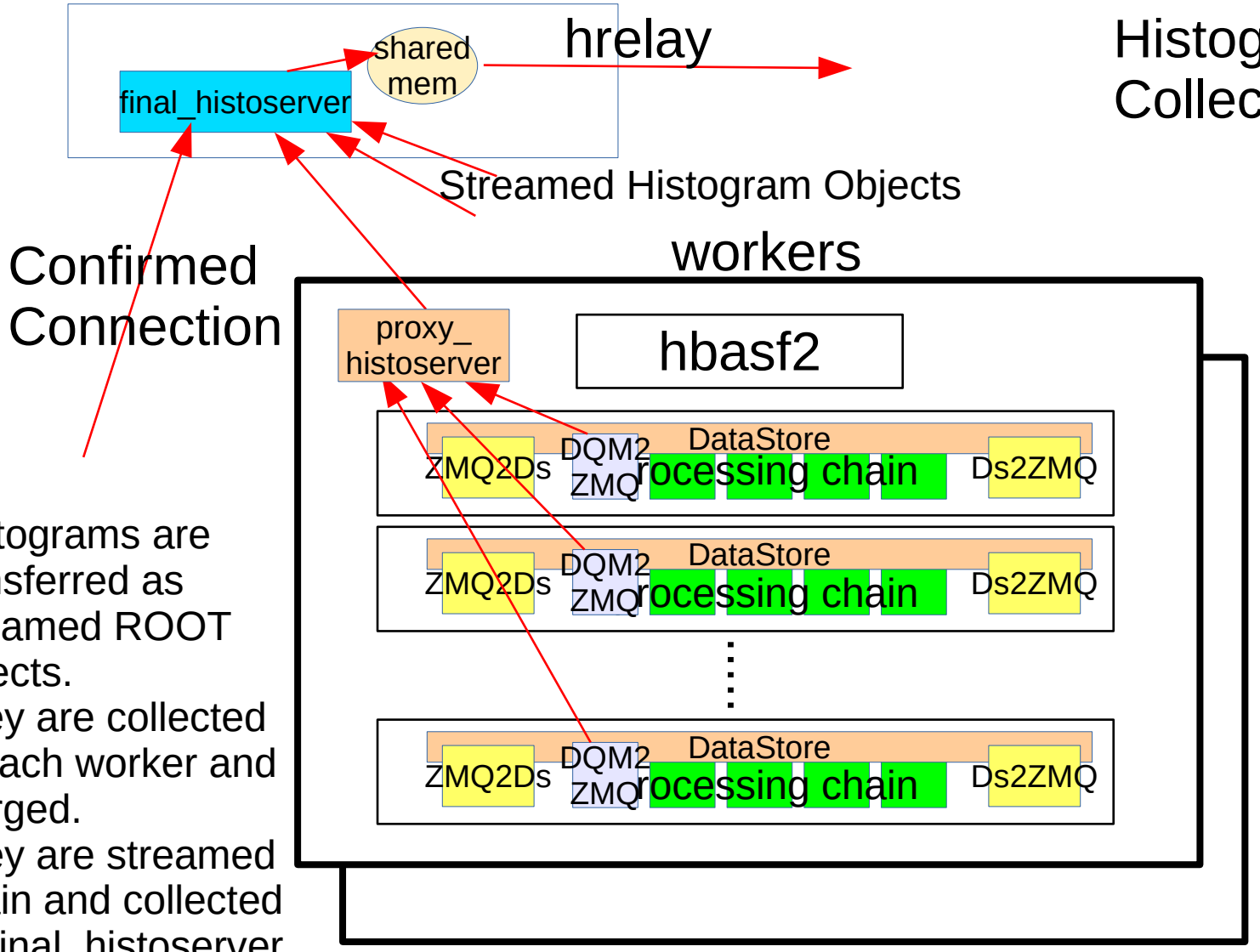
- RoI binary is embedded in ZeroMQ message as a separate packet.

- System control is integrated in the Belle II standard slow control package (daq_slc).

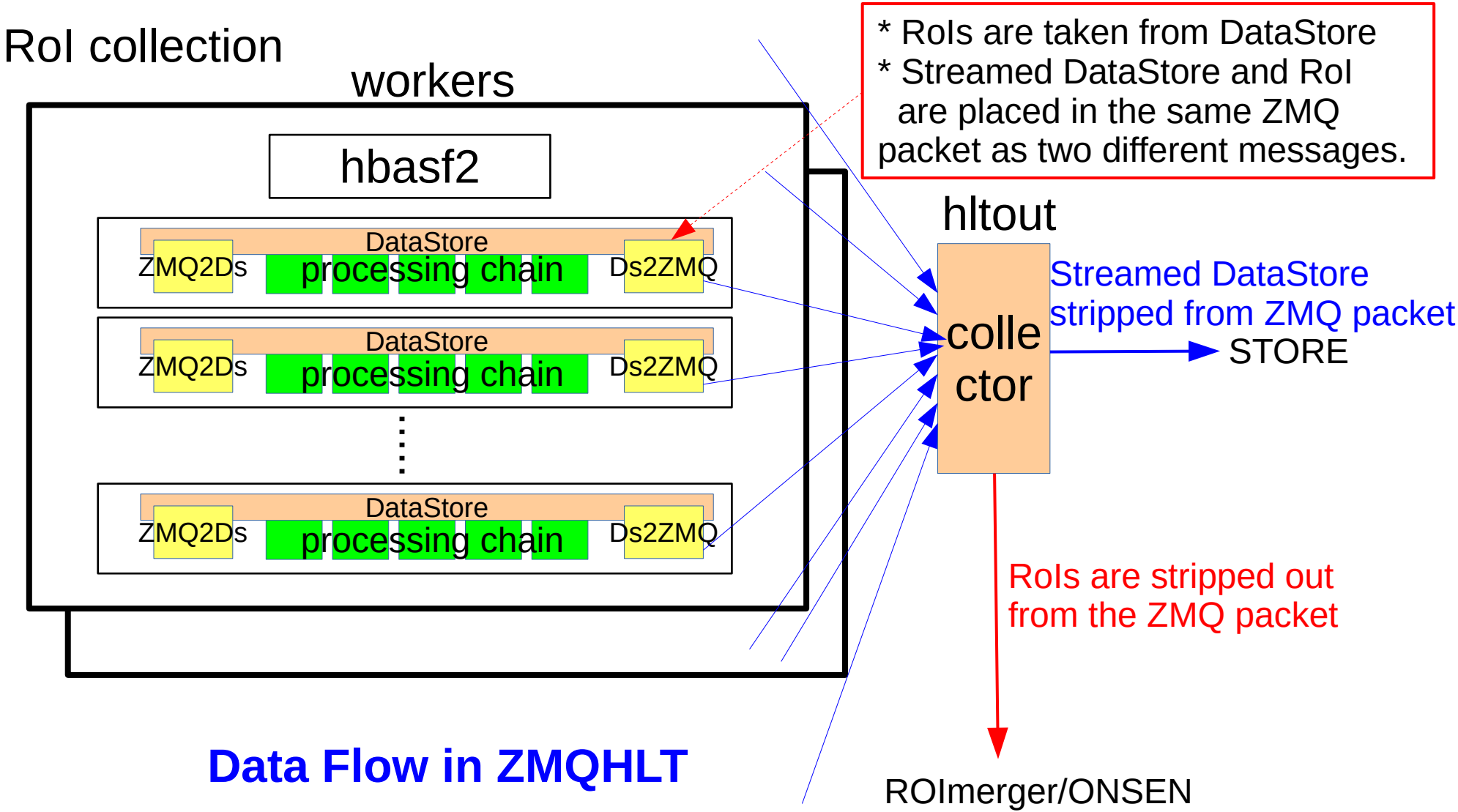


Data Flow in ZMQHLT

Histogram Collection with ZMQ



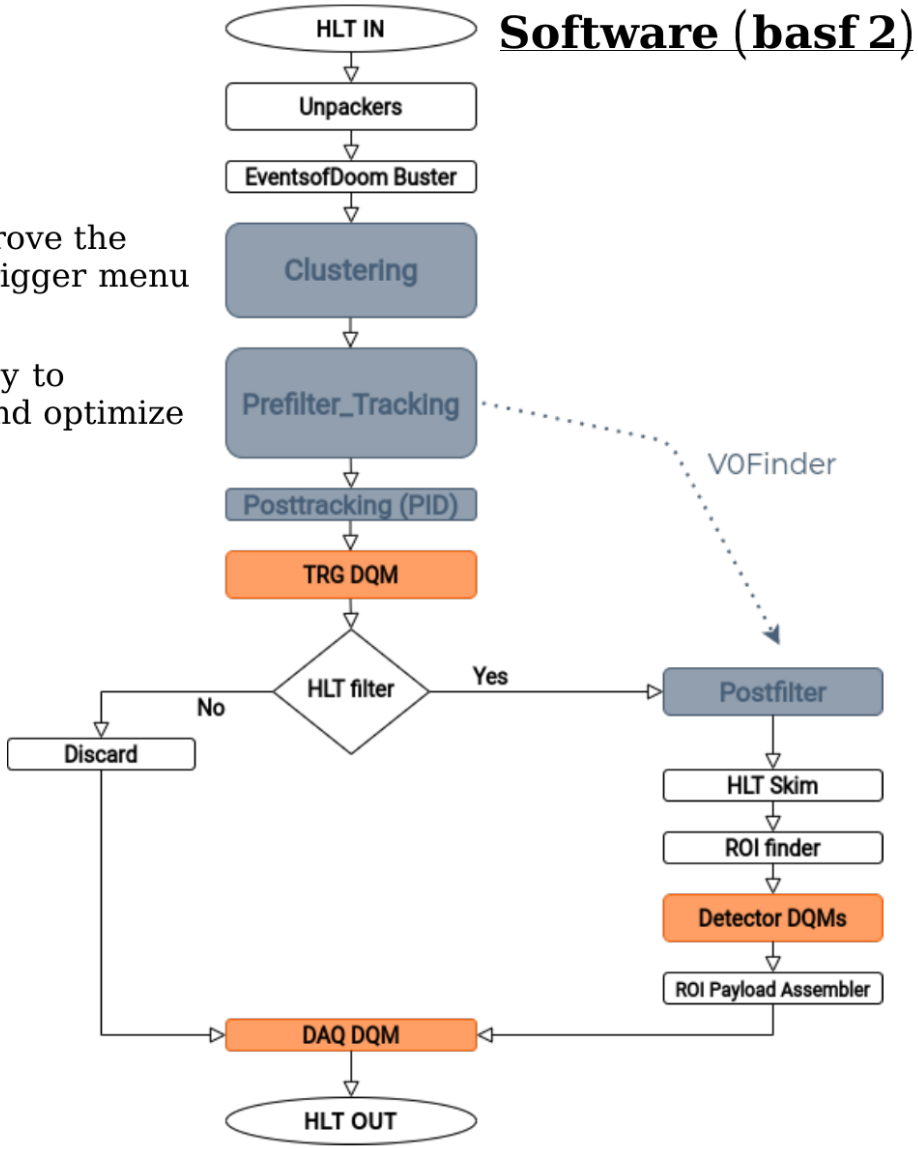
- Histograms are transferred as streamed ROOT objects.
- They are collected in each worker and merged.
- They are streamed again and collected by final_histoserver.


RoI collection
workers
hbasf2
DataStore
ZMQ2Ds
processing chain
Ds2ZMQ
DataStore
ZMQ2Ds
processing chain
Ds2ZMQ
DataStore
ZMQ2Ds
processing chain
Ds2ZMQ
* RoIs are taken from DataStore
* Streamed DataStore and RoI are placed in the same ZMQ packet as two different messages.
hltout
colle
ctor
Streamed DataStore
stripped from ZMQ packet
STORE
RoIs are stripped out
from the ZMQ packet
ROImerger/ONSEN
Data Flow in ZMQHLT

HLT is software

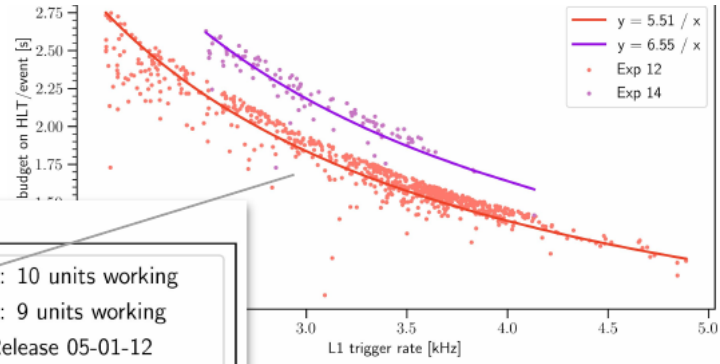
Software optimization is the key to improve the situation in addition to tightening L1 trigger menu

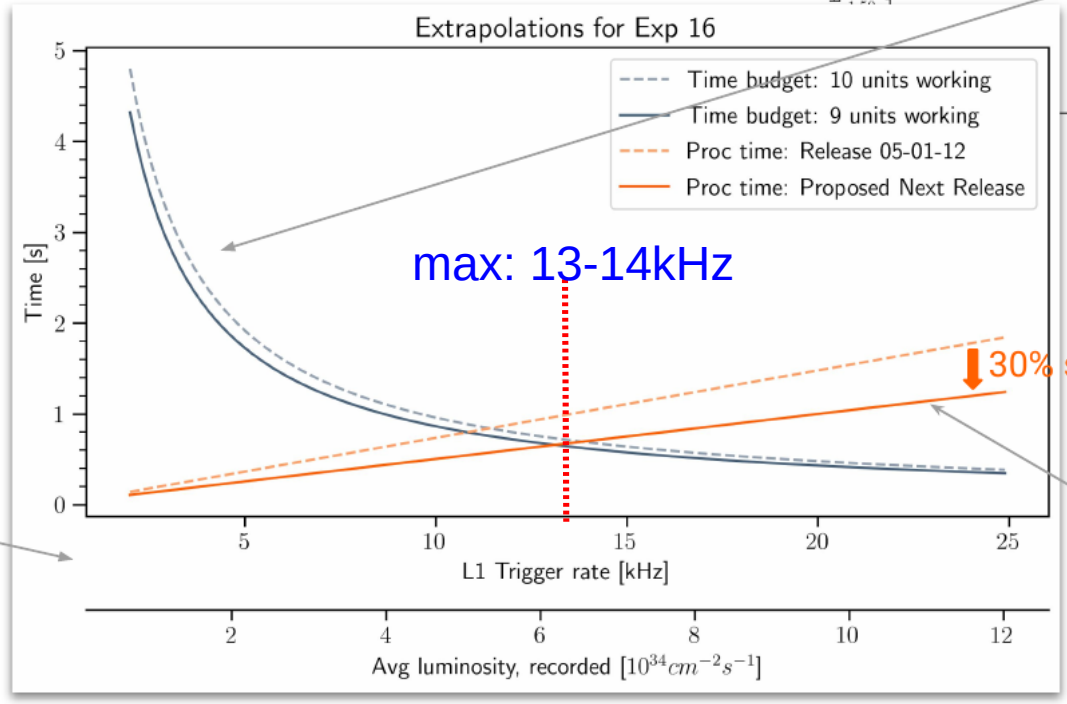
Monitoring during data taking is the key to understand better the HLT operation and optimize its performances (CPU and memory)

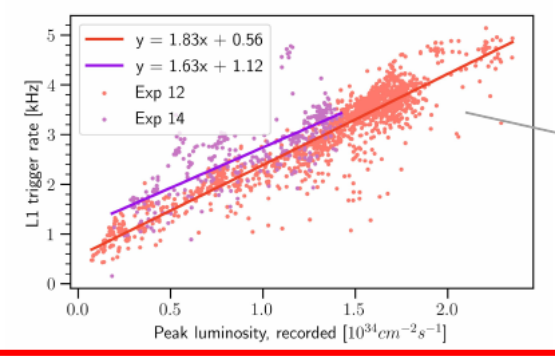


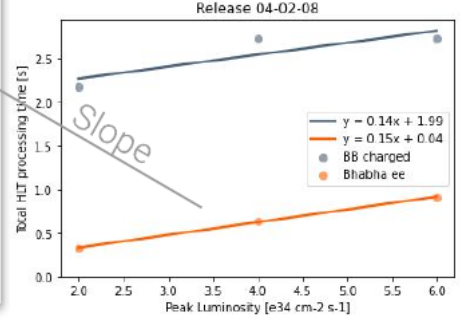
Estimates for 2021 a/b

- 2020 winter upgrade is included
- 9 out of 10 units working
- # processes is full, i.e., no out of memory problem
- No hardware under-performance (10% observed in Exp 14)









If the L1 trigger menu stays the same.

Should be safe for 2021 a/b

Generated samples at different luminosity (background conditions) to extrapolate processing time growth

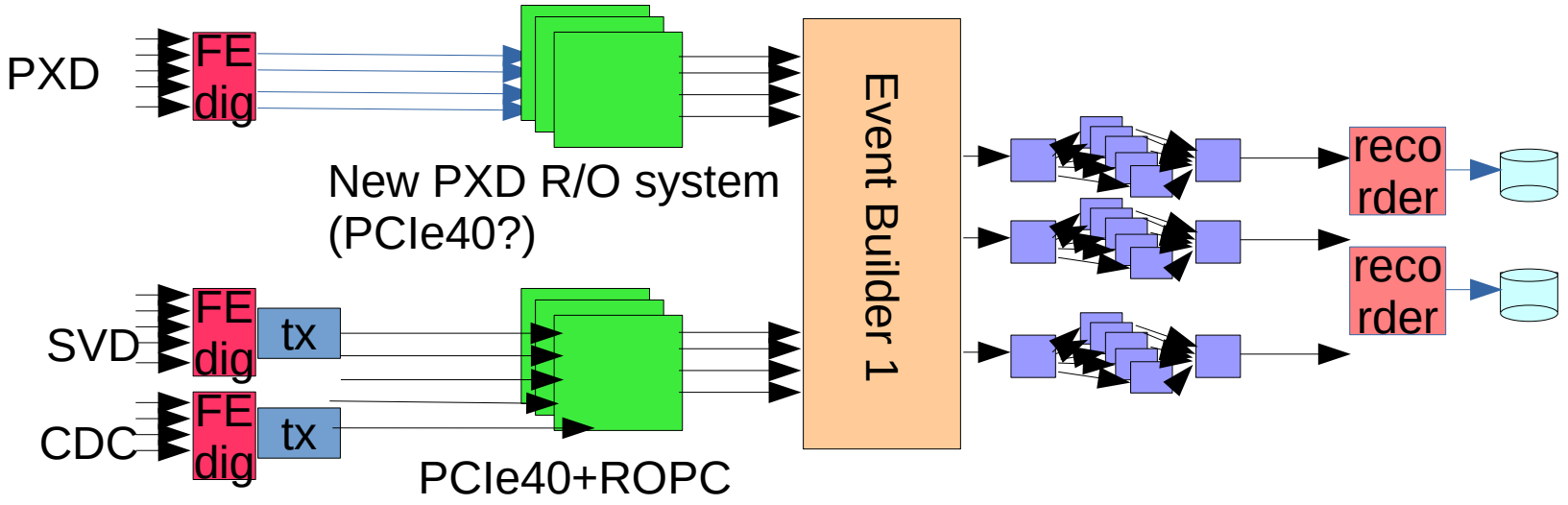
3. Some thoughts on slow pion tracking with PXD in Belle II DAQ

- Goal: Implement a mechanism to rescue (super low momentum) slow pion associate hits in PXD by the fast tracking with PXD+SVD.

- Three possible options:
  a) Full HLT software processing by feeding PXD data into HLT w/o reduction.
  b) Hardware tracking with PXD+SVD in PXD readout to recover slow pion hits in RoI selection. PXD data feed into HLT.
  c) b) + RoI feedback from HLT. No PXD data feed into HLT.

- Performance assumption
  * PXD data size w/o reduction : ~1MB for 3% occupancy
  * 30kHz maximum

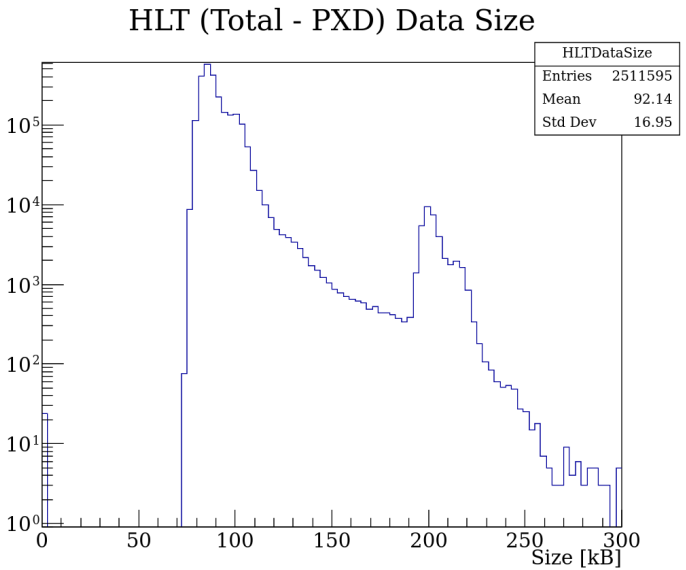
a) Full software processing



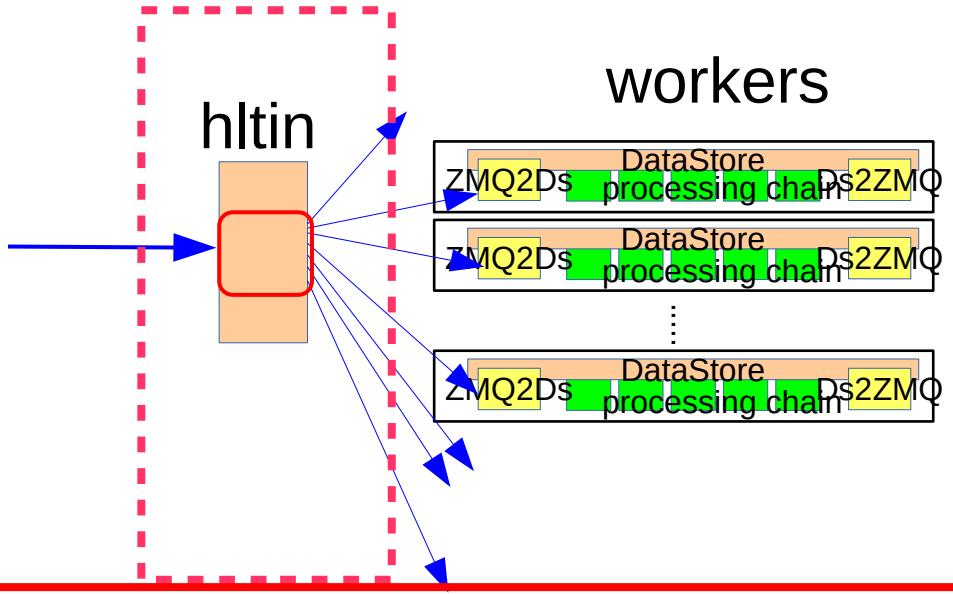
- In this case, HLT needs to handle 10 times larger input data flow.
  <- no data reduction for PXD at the input of HLT..... (i.e. no RoI reduction)
  * 1.2MB/event * 30kHz. <-> ~200KB/event * 30kHz
  * Large asymmetry in data flow to event builder 1.
  * Larger memory size needed in each HLT workers.

➡ Not feasible!

Current event size

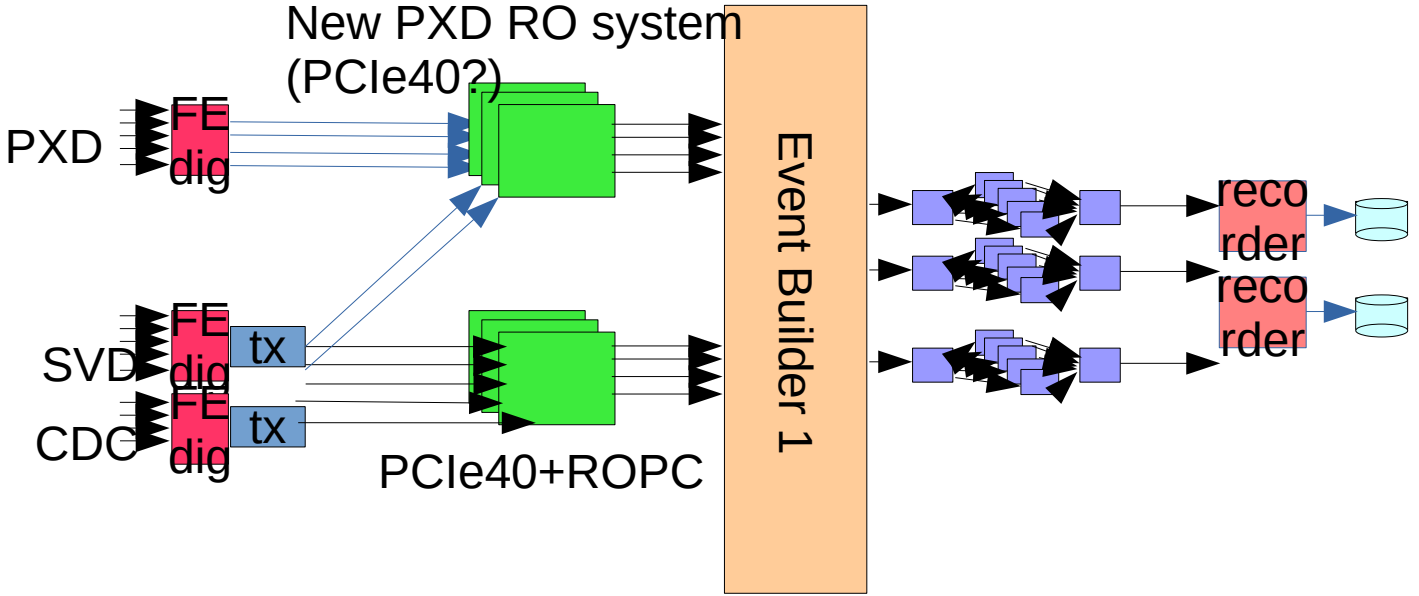


If PXD data flow is fed into HLT w/o reduction:
with 10 HLT units : (1+0.2)MB * 30kHz / 10 => 3.6GB/sec/unit



- Event distribution is currently handled by 10GbE network
- It cannot handle 3.6GB/sec data flow.
  => Need the upgrade to 100GbE network, but it costs much.
- Current memory size of each worker is 2GB/core. The usage is already > 80% in the current (bad) beam condition.
  => Need to have 10 times larger memory to accept the data flow.
- Additional processing power for slow pion tracking.
  => More CPU required.

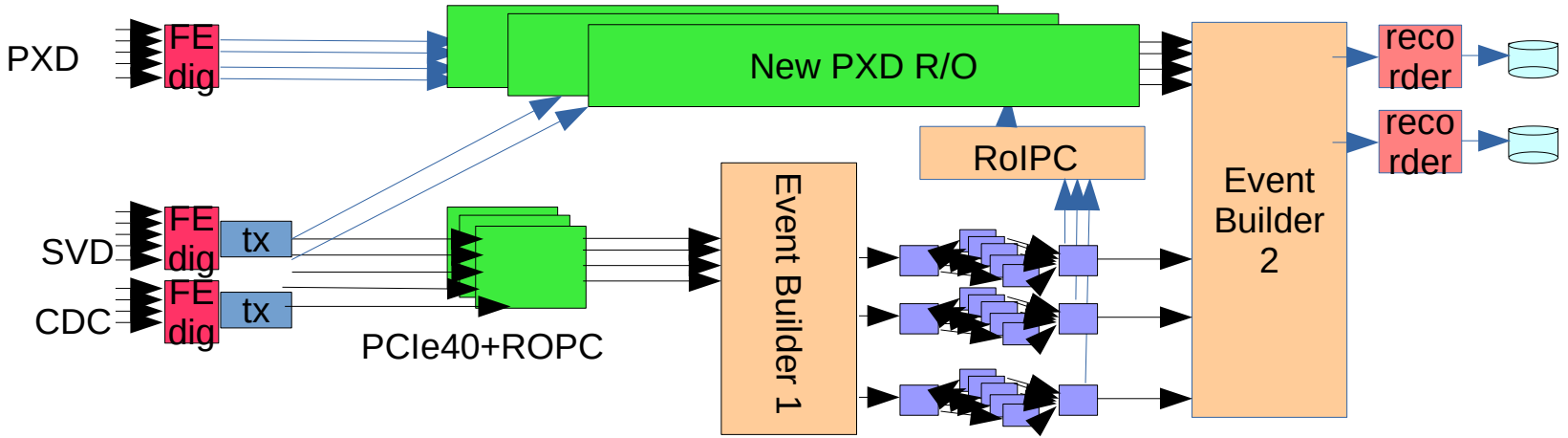
b) Hardware Tracking + ROI selection in new PXD RO



- Original idea of PXD readout by Christian Kiesling san.
- If the event size reduction of 1/10 by RoI using hardware tracking in PXD RO system is ensured, the scheme is feasible with some reinforcement in HLT processing power.

- However, the hardware tracking (DATCON) is not yet established, and a bit dangerous since there is a possibility that RoI reduction does not work.

c) b) with HLT RoI

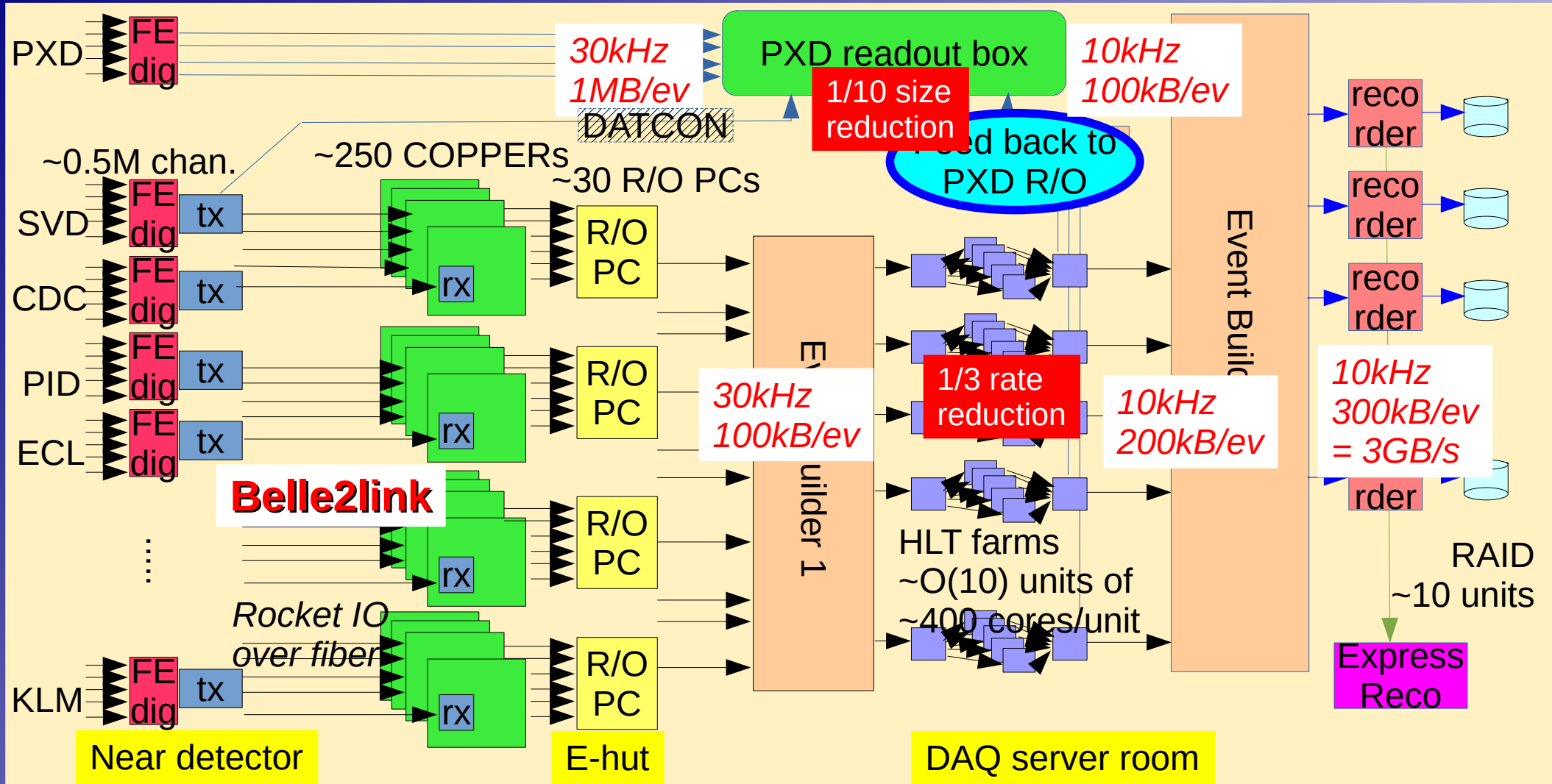


- Basically the concept is the same as b) but keep the HLT RoI reduction.
- In new PXD readout, SVD+PXD hardware tracking is performed and the PXD hits associated with slow pion tracks are kept after RoI selection.

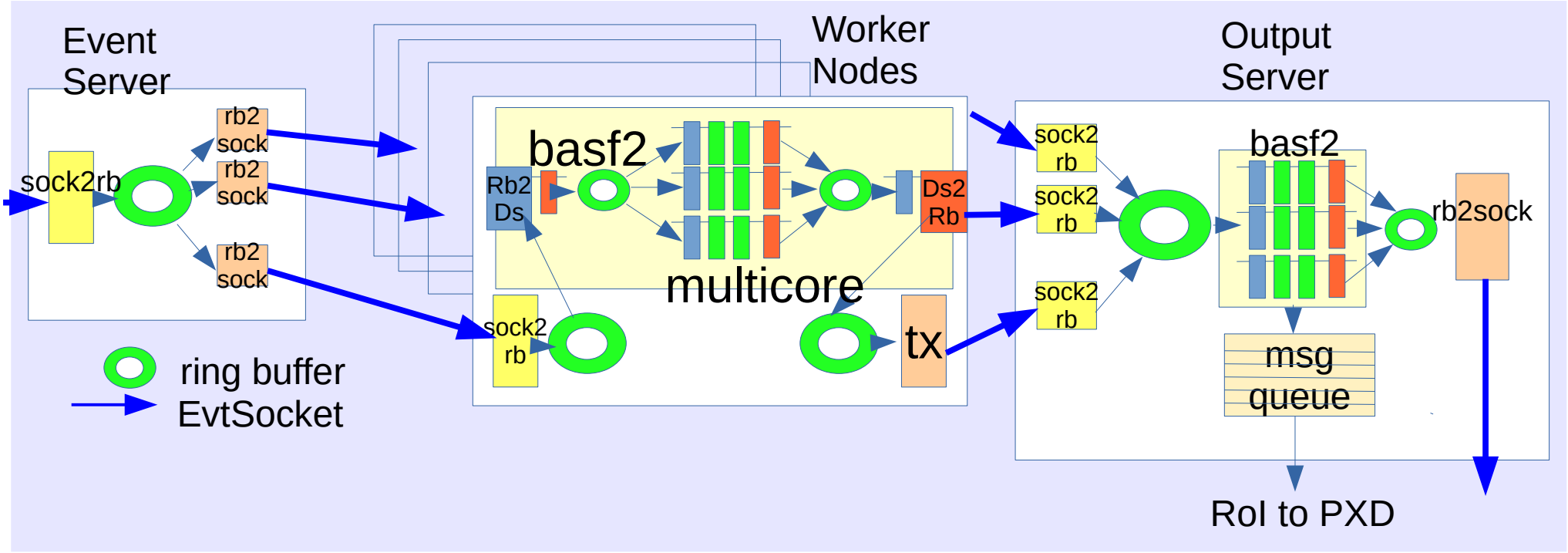
- The demerit of this approach is that the rescued slow pion hits cannot be utilized in HLT selection, since PXD data are not fed into HLT.
  => However, slow pion hits are kept in PXD raw data.
  Do we need super low momentum slow pions for HLT decision?
  If yes, how much improvement (in physics) is expected?

Backup Slides

Data Flow in Belle II DAQ



HLTData Flow



- The event data from eb1 are distributed to HLT processing nodes via socket connection + RingBuffer.
- They are “objectized” at each core or HLT nodes.
- The output are streamed as “EvtMessage” and collected via socket connection + RingBuffer => switched to ZeroMQ.

1. HLT Reinforcement : Hardware

- In 2019, HLT was operated with
  * 9 HLT units with ~320 cores/each. -> 2880 cores in total
    (Note: the number of “physical” cores)

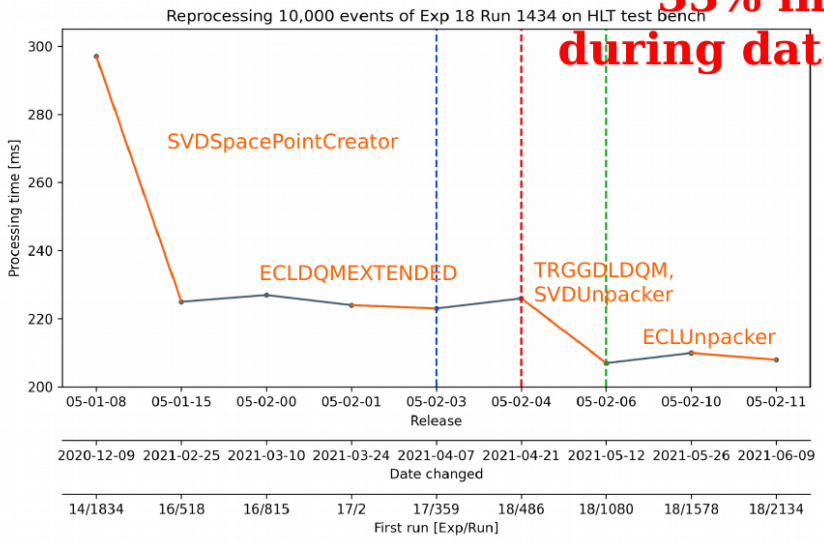
- Estimated maximum processing rate ~ 8kHz.

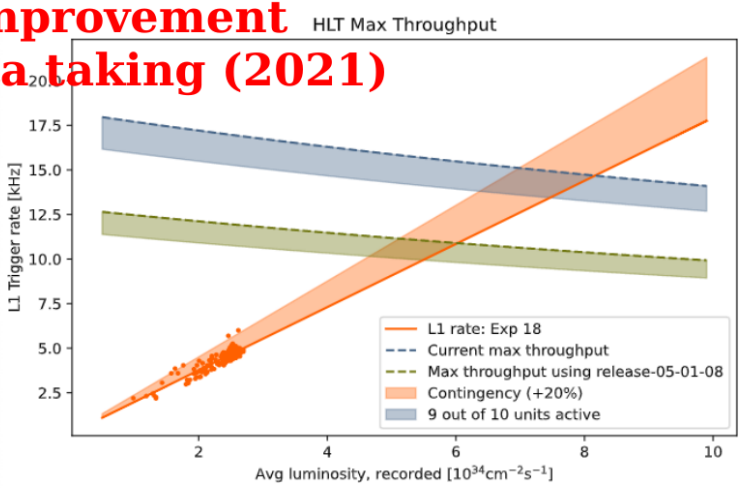
- Maximum L1 rate was 3.5kHz @ $L = 2 \times 10^{34}$

Should be prepared for higher rate > 10kHz for 2020-2021 run.

Optimizations implemented (new releases) during 2021

33% improvement during data taking (2021)





~ 9 kHz → ~ 13 kHz

Should now be able to operate HLT smoothly till LS1 !

Further improvements for release-06 (including efforts from TOP/SVD): ~−15%

HLT Upgrade Timeline

Full design : 6400 physical cores