Commissioning and Operation of the Upgraded Belle II DAQ System with PCI-Express-Based High-Speed Readout

Yun-Tsung Lai

on behalf of the Belle II DAQ upgrade group

KEK IPNS

ytlai@post.kek.jp

計測システム研究会 2022 @ J-PARC

18th Nov., 2022

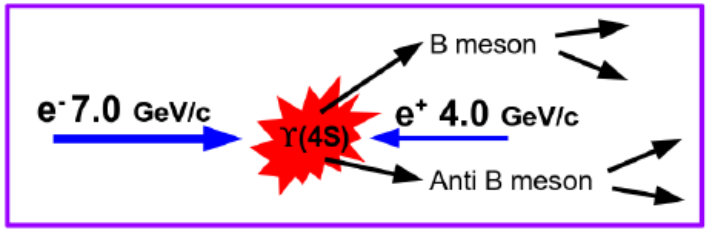
Outline

- Introduction
- Readout system in Belle II DAQ: COPPER and PCIe40
- Development of the PCIe40 readout system
- Upgrade
- Performance
- Summary

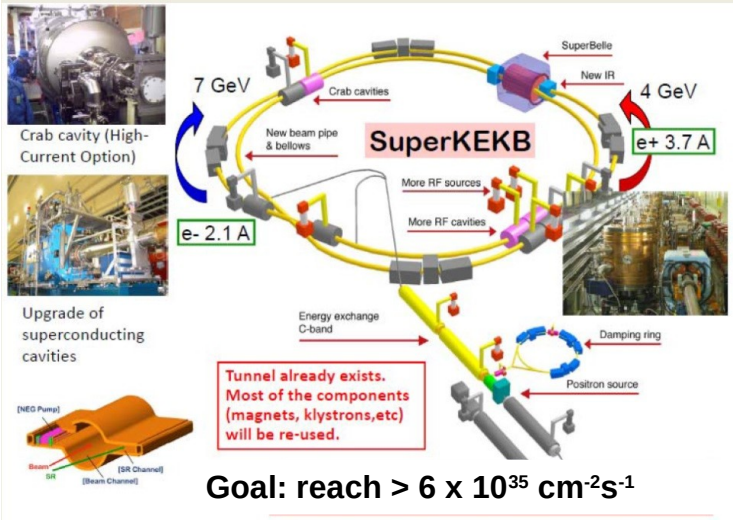
SuperKEKB

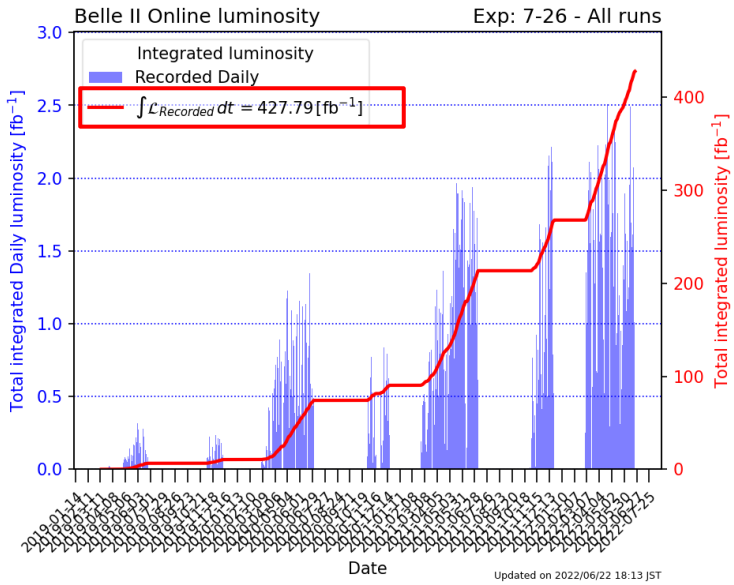
- SuperKEKB: Upgraded from KEKB.
  - More than 30 times larger luminosity of KEKB with nano beam scheme.

- Asymmetric energy collider:
  - 7.0 GeV e^- and 4.0 GeV e^+ for $\Upsilon(4S) \rightarrow B\overline{B}$.



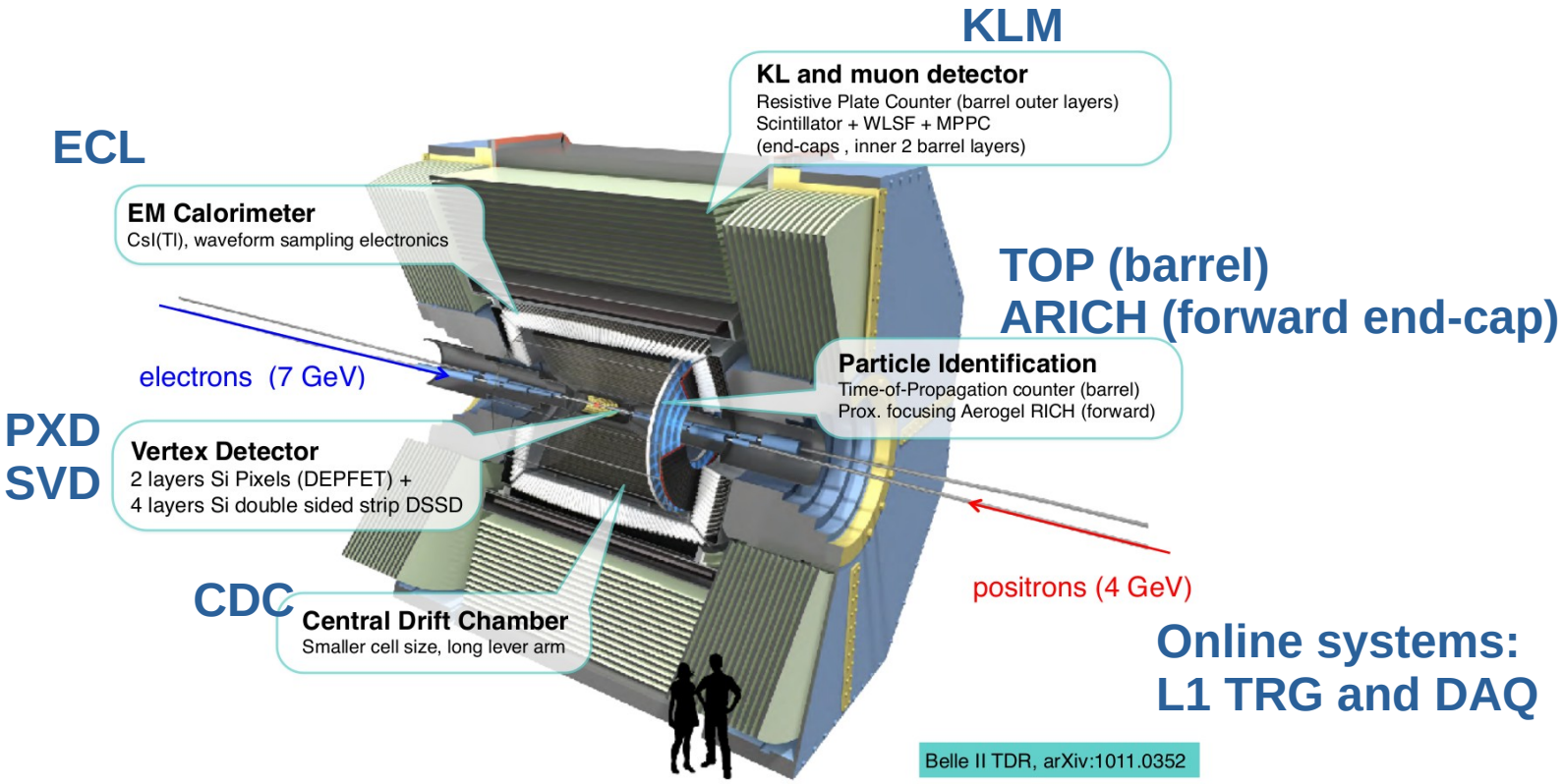
- Luminosity achievement:
  - L_{peak} = 4.65 x 10^{34} $cm^{-2}s^{-1}$.
    World record. ~Two times of KEKB record with much smaller beam current.
  - L_{int} = ~427 fb^{-1} up to Jun. 2022.





Belle II detector

- Belle II: Newly-designed sub-detectors set to improve detection performance.

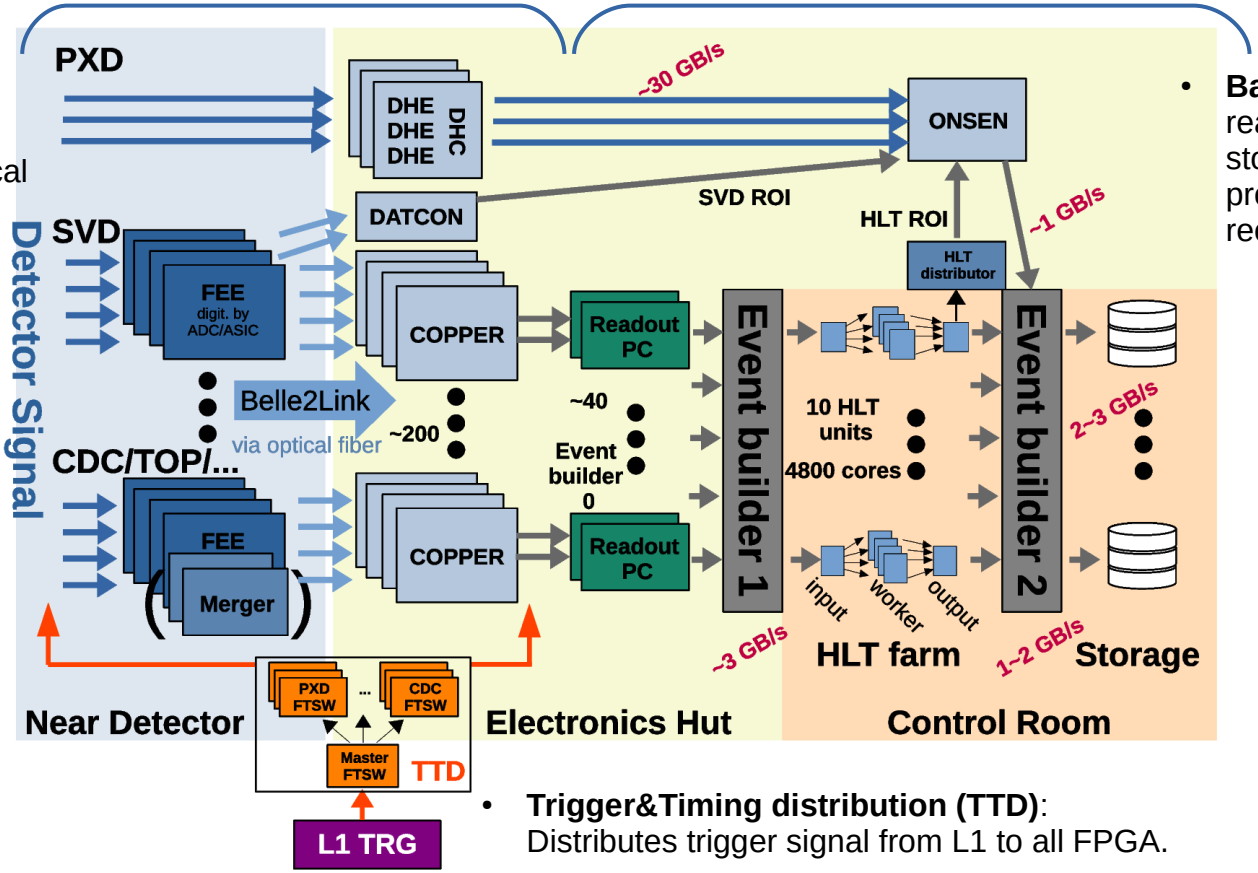


- Physics target of Belle II:
  - Rare B, τ, charm physics, Dark Matter search, CP Violation.

Belle II DAQ system

- Pipeline readout system with common design for most of the sub-detectors.
  - Except for PXD, which has a data reduction system with ROI.
- Target of performance: 30 kHz trigger rate from L1, ~1% of dead time, and a raw event size of 1 MB.

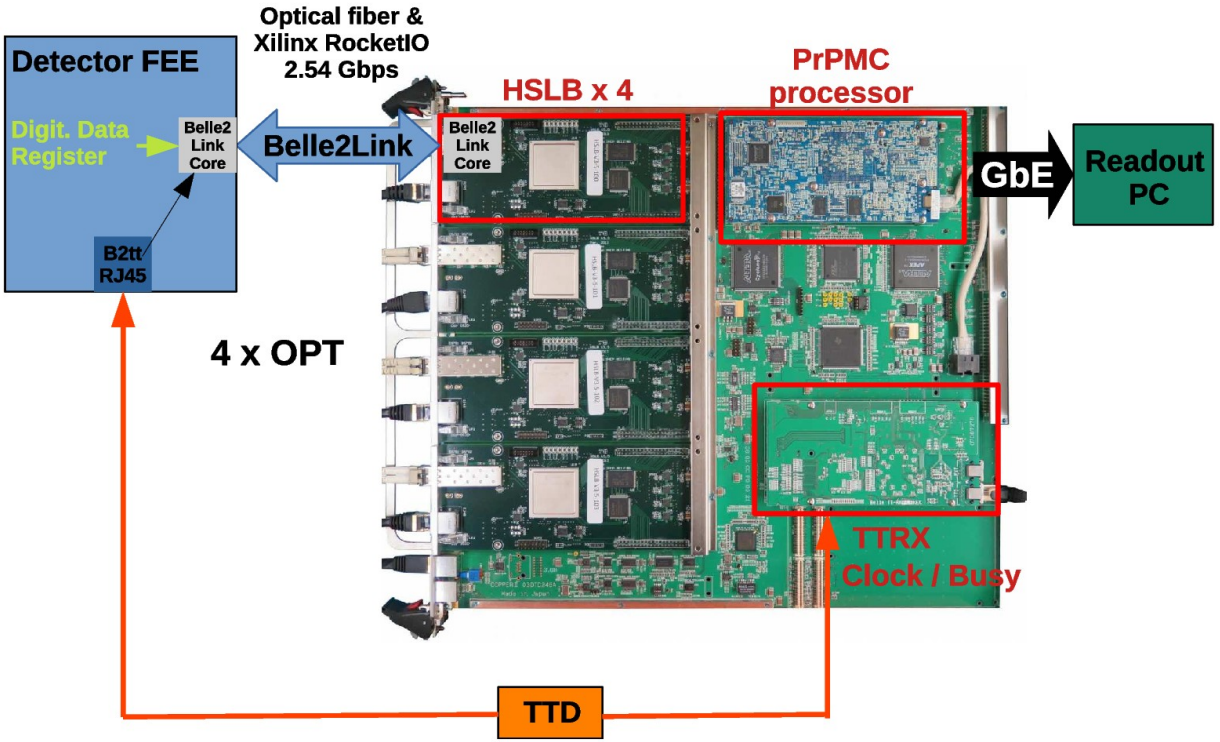
- **FPGA (FEE - Readout)**: Use universal "Belle2Link" protocol with optical links in between.



- **Back-end servers**: readout PC, HLT and storage for online procession/ reconstruction.

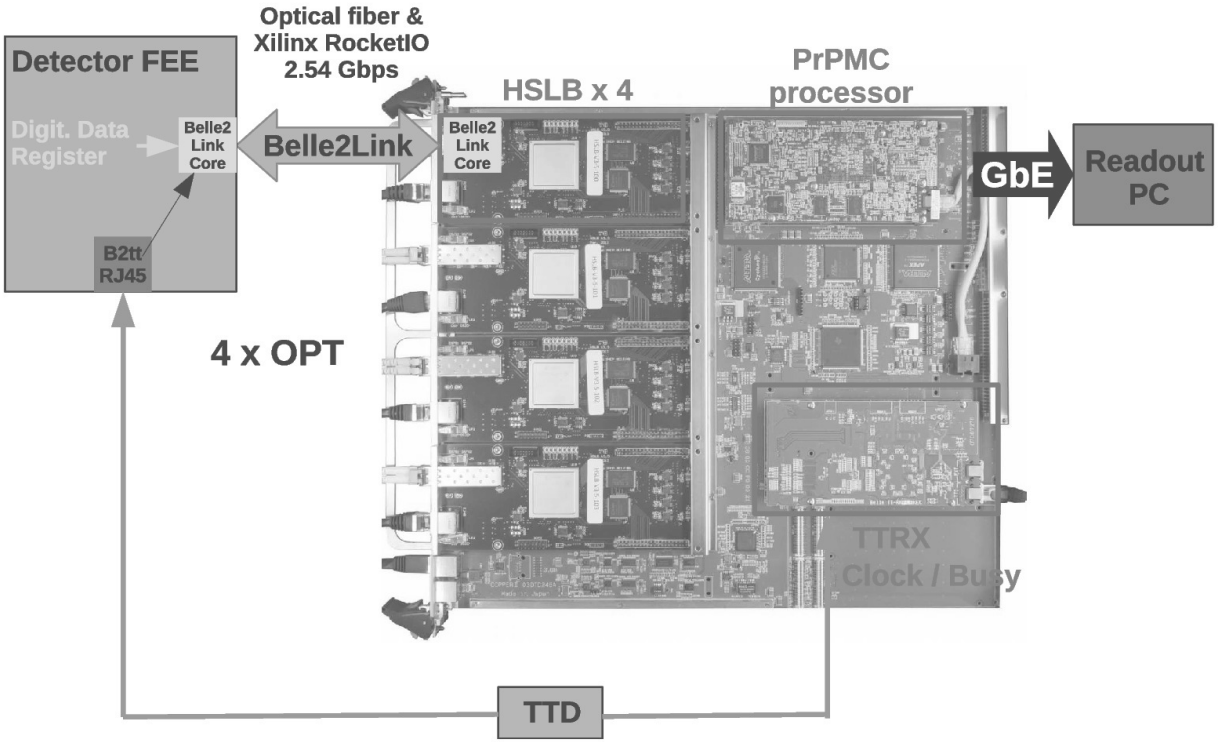
- **Trigger&Timing distribution (TTD)**: Distributes trigger signal from L1 to all FPGA.

COPPER readout



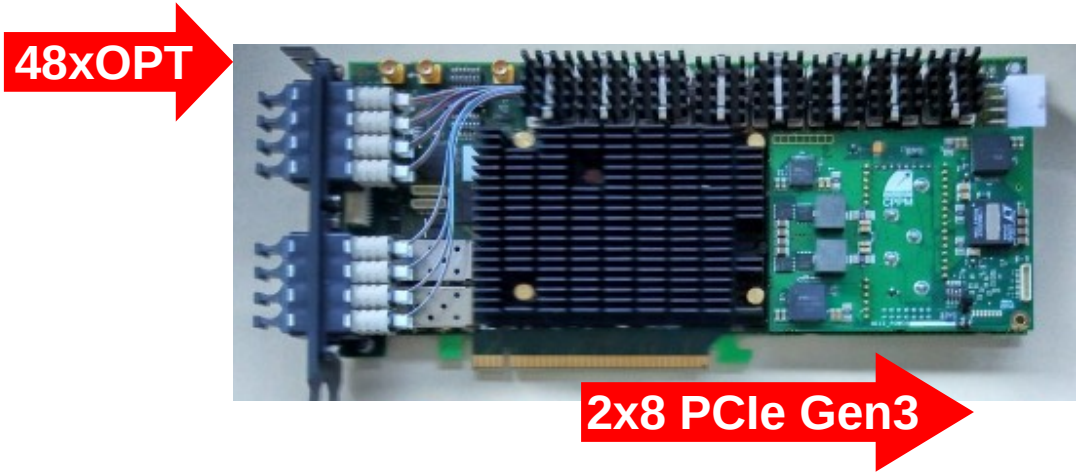
- **COPPER board**:
  - From the era of Belle.
  - 9U VME.
  - 4 Xilinx Virtex-5 receiver boards (HSLB): receive data from 4 detector FEE.
  - PrPMC: data procession, pre event building.
  - TTRX: interface to TTD system.
  - In total 203 COPPER boards were used in Belle II.

Considerations for readout upgrade in Belle II DAQ



- **Difficulty of maintenance**:
  - 203 COPPER x 6 daughter boards: Large amount of separated pieces.
  - Increasing number of malfunctioning pieces since Belle II operation.
  - Parts out of production already.
- **Limit of the system on further improvement:**
  - Output throughput by GbE: 1Gbps.
  - CPU usage of PrPMC: ~60% at 30 kHz trigger rate.

Upgrade: PCIe40 board with new readout PC



- **PCIe40 board:**
  - Based on an Intel Arria 10 FPGA.
  - Has been developed in upgrade of LHCb and ALICE experiments. Belle II has its own firmware/software design.
  - 48 transceivers for optical links from 48 FEE.
  - 2x8 PCIe Gen3 to a readout PC (1-to-1 connection).
  - In total 21 sets of PCIe40 + readout PC will be used in Belle II.

Development items

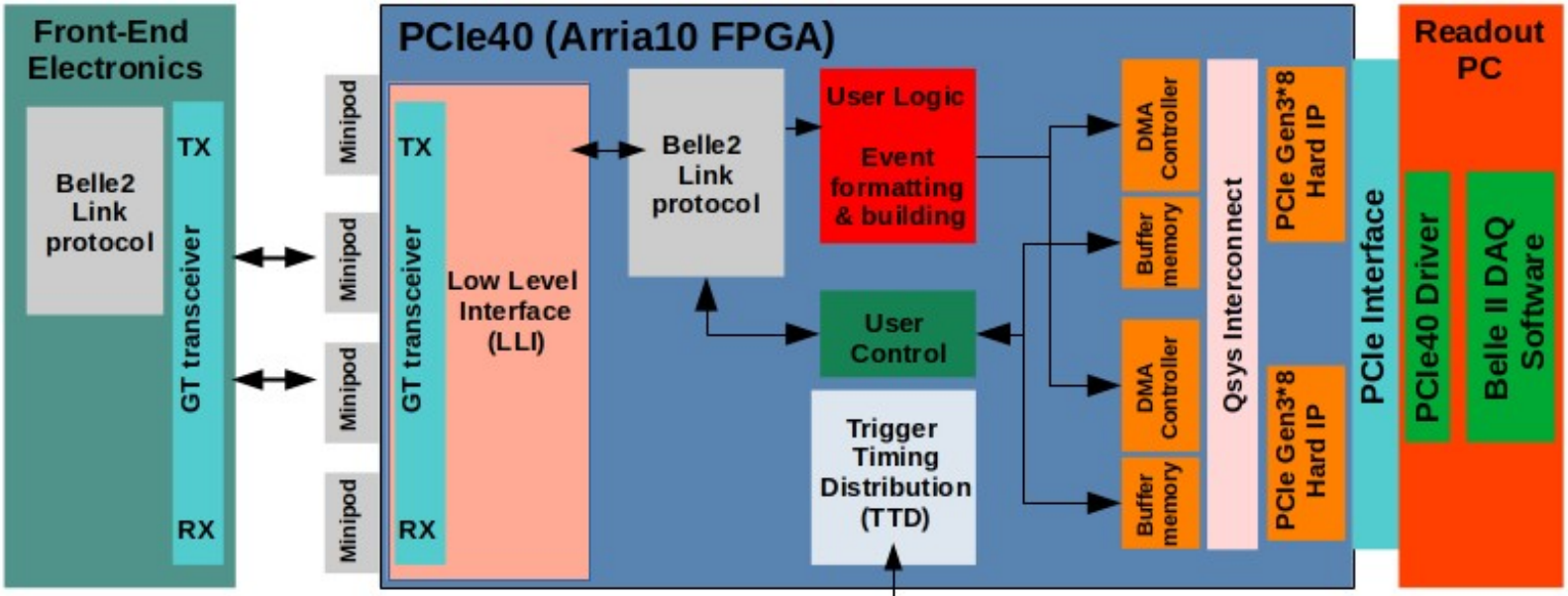
Belle2Link protocol:
8B/10B with transceiver & optical link.

PCIe40 firmware

DMA for PCIe:
Interface between PCIe40 and readout PC.

Slow control: register rd/wr with B2Link
Detector configuration & System monitoring.

Data flow of detector FEE data. Processed inside PCIe40.



- Interface to other system: FEE, TTD, readout PC.
- Data procession logic for formatting and first-level building.
- Changes on other systems should be minimal.

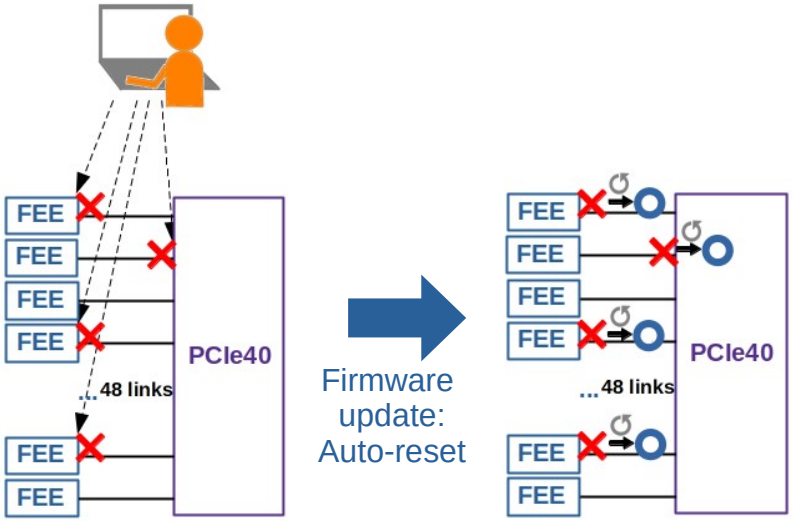
To TTD system via FTSW:
Trigger & timing distribution.
Masking, run control.

Software in readout PC:
- PCIe readout
- Event building
- Slow control

Automatic link recovery

- A PCIe40 is connected to 48 FEE with 48 links
  - Once it is reprogrammed, those 48 links are affected, and ~50% of them could be down.
  - To recover them, we need to reprogram the corresponding FEE and repeat.
    - Manual recovery via CUI/GUI: time-consuming.

- **Improvement: Auto-reset on the link inside firmware.**
  - Monitoring status flags: PLL lock, decoding error, disparity, etc.
  - Once instability is observed, issue reset signal to the transceiver automatically.
  - Reliable recovery: ~100% readiness.
    Manual recovery is not needed.
  - Implemented in PCIe40 and all FEE firmware based on different transceivers:



Detector FEE	Transceiver
SVD	Spartan-6 GTP
CDC	Virtex-5 GTP
TOP	Kintex-7 GTX
ARICH	Virtex-5 GTP
ECL	Spartan-6 GTP
KLM	Virtex-6 GTX
TRG	UT3: Virtex-6 GTX, GTH UT4: UltraScale GTH, GTY

Universal Belle2Link protocol

Development by IHEP:
D. Sun et al., Phys. Procedia, vol. 37, pp. 1933-1939, 2012.

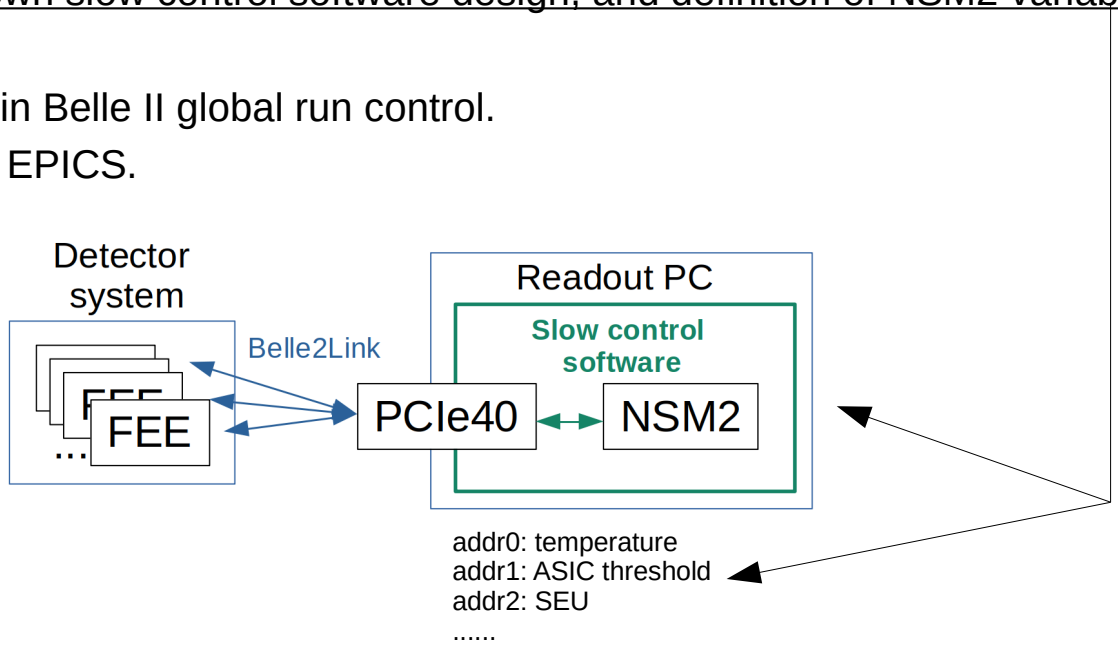
- **Belle2Link protocol**: universal for each detector FEE - readout (COPPER/PCIe40)
  - Line rate 2.54 Gbps.
  - Framing transmission using different 8B/10B K characters.
  - The same protocol design is implemented in PCIe40.
  - Using optical link and high-speed transceivers of FEE and PCIe40.
  - Two major functionalities:
    - Transferring **detector FEE data** w.r.t L1 trigger.
      crc16 and crc32 checksum included.
    - **Slow control**: exchanging register content between FEE and readout as a set of address and payload data.
      - Each detector system has its own slow control software defining the meaning of each address and payload.

Slow control software

- Belle2Link: FEE and PCIe40 exchanges information as address/data.

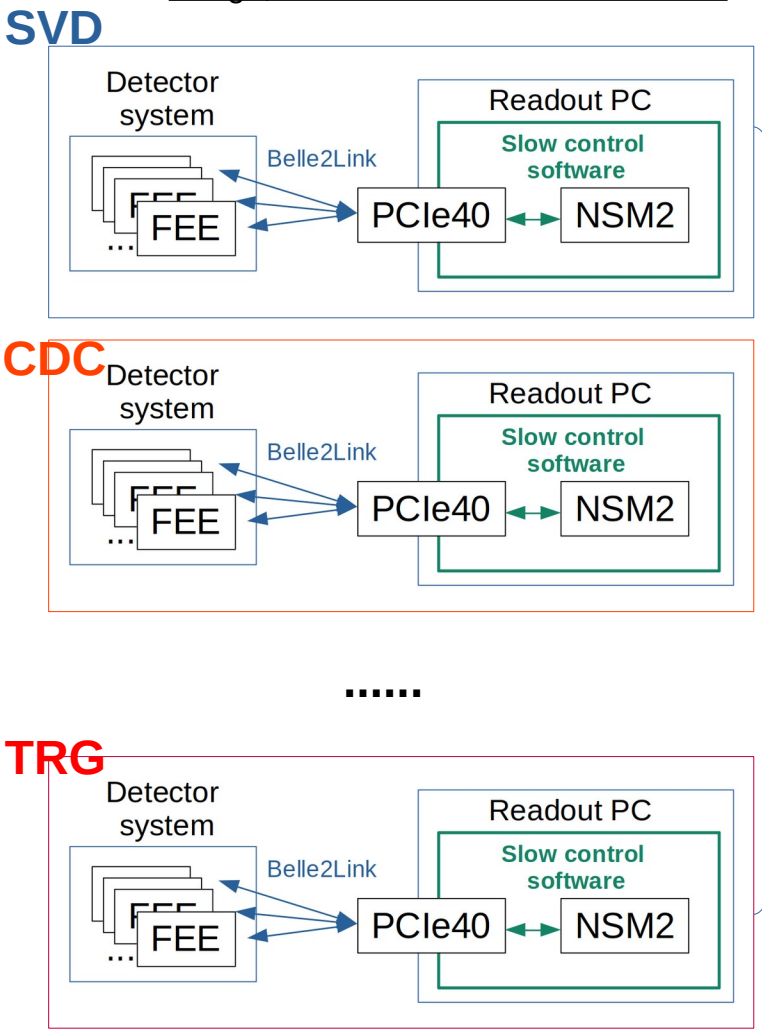
- Slow control software:
  - Runs in readout PC, and controls Belle2Link.
  - NSM2: Network Shared Memory v.2.
    Define address/data as variables.
  - Configuration and monitoring for each detector.
  - **Major development item in the PCIe40 upgrade project**: Each detector system has its own slow control software design, and definition of NSM2 variable.

- Integrated in Belle II global run control.
- Logged by EPICS.

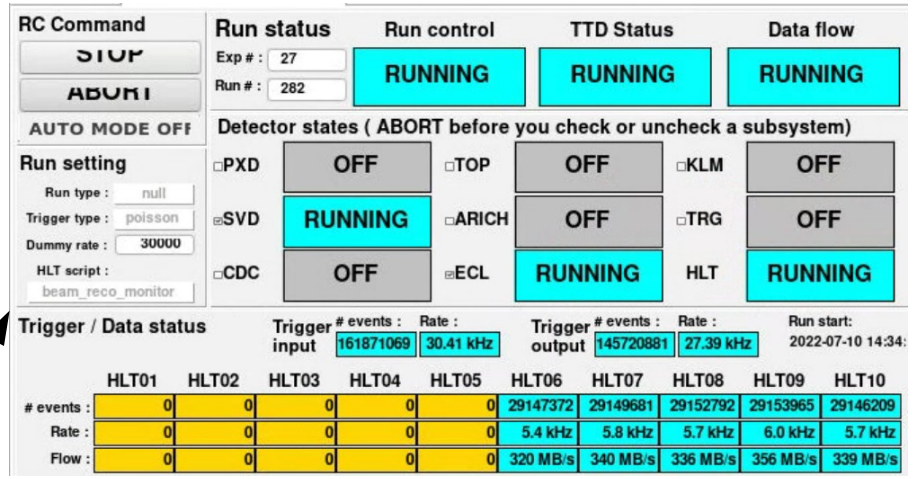


Slow control of Belle II DAQ

Major development: Each detector system has its own slow control software design, and definition of NSM2 variable

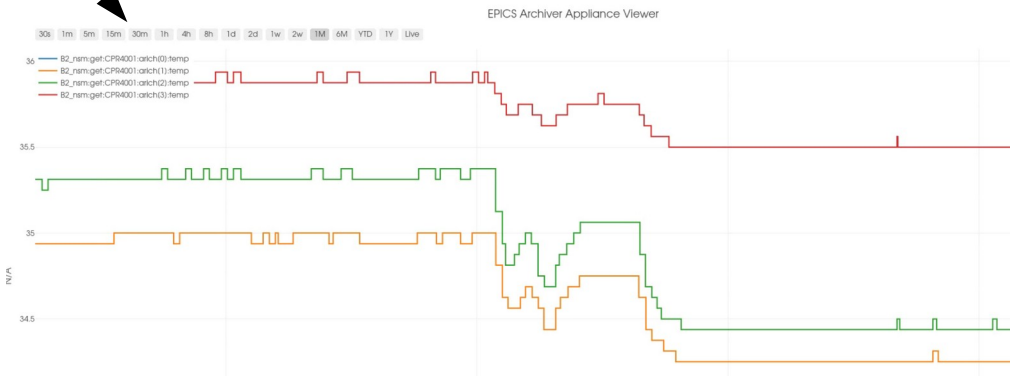


GUI for Belle II global run control



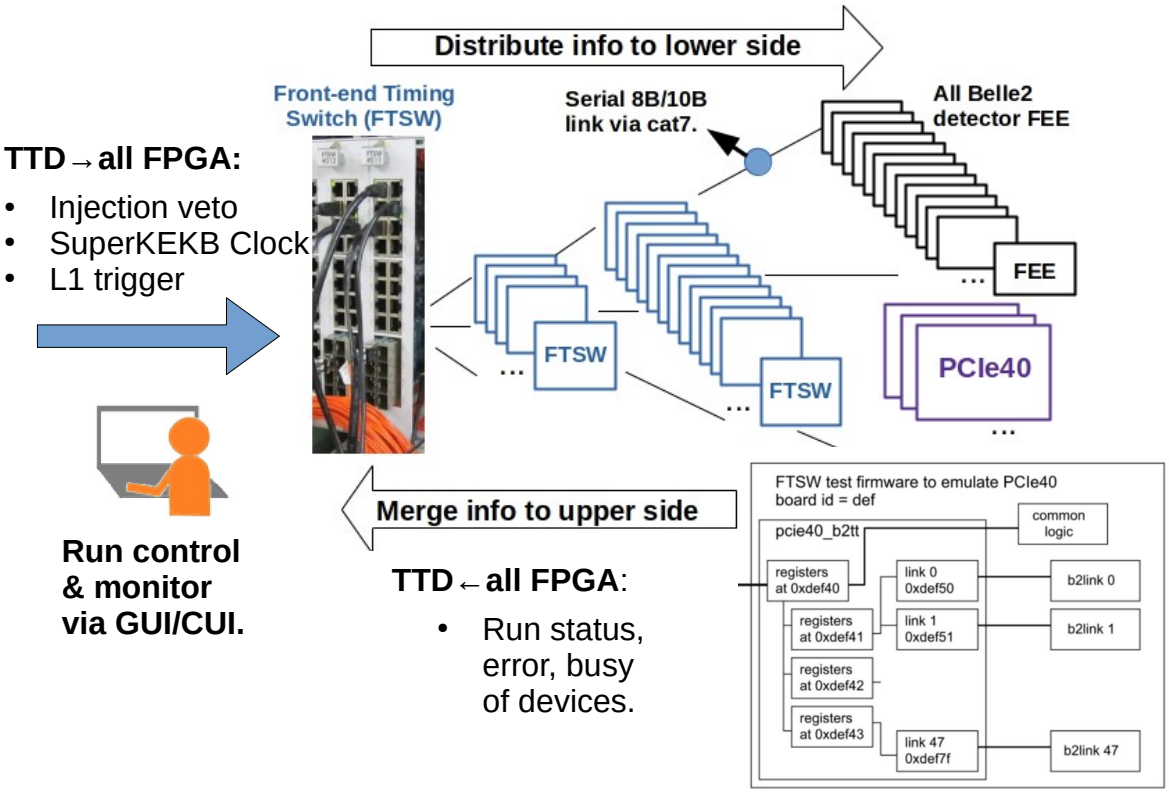
EPICS

Plots of archived information



Interface to the TTD system

- Trigger&Timing Distribution (TTD) system:
  - Distributing/Collecting information to/from FPGA devices (FEE and readout) in real-time.
  - Using Front-End Timing Switch (FTSW) boards.
  - Run control purpose.
  - Adaptation in PCIe40 firmware design.

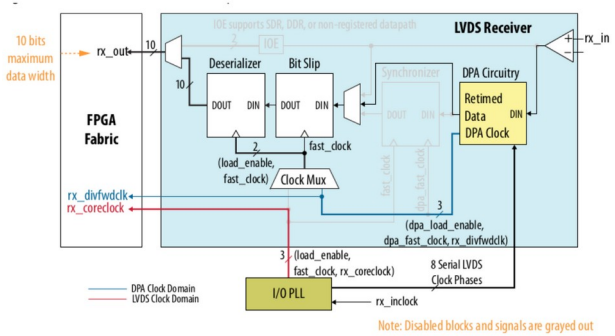


- PCIe40: Information of all 48 channels needs to be reported:
  - New address scheme to merge 48x info is developed in PCIe40 firmware, such that FTSW firmware is not touched.

Interface to the TTD system: Improvement

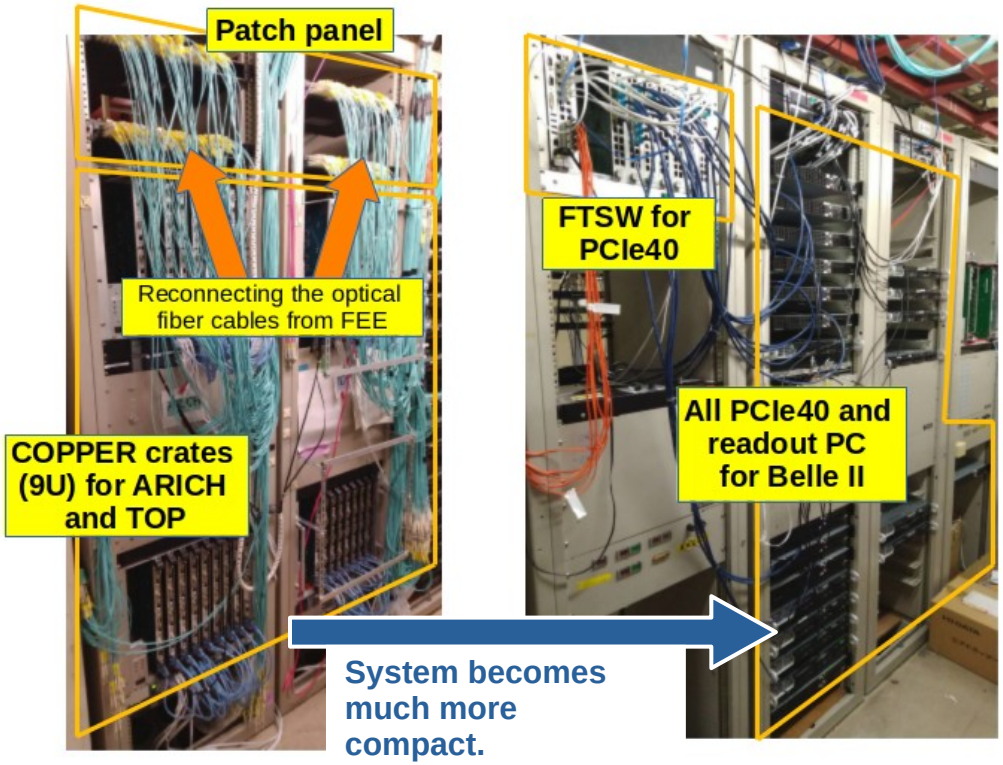
- Clock, and the signals (to be driven by the same clock source) are distributed by TTD system to PCIe40:
  - Stability affected by external noise in Electronic Hut.
  - TTD link down sometimes happens: down time in operation.

- **Improvement: Using on-board clock with Intel Serdes IPcore.**
  - On-board clock: Stable under external noise.
  - Soft-CDR to handle jitter.
  - Reduce operation down time.



- **Future plan: Replace cat-7 cable with optical fiber cable.**
  - By some investigations, it provides better stability in transmission.
  - Firmware is under development, and will be complete in LS1.

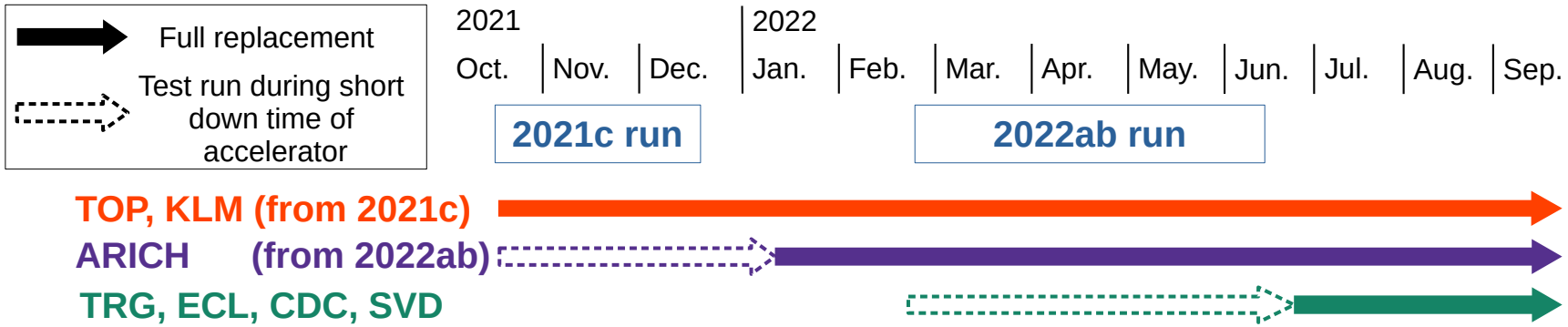
New system in Belle II Electronics Hit



- In Belle II Electronics Hut:
  - Patch panel for optical cable is prepared to change the FEE connection from COPPER to PCIe40.

- A set of PCIe40 + it host readout PC occupies only 1 slot of a rack.
  - Copper system: Lots of 9U crates,
  - The new system is much more compact.

History of PCIe40 upgrade in Belle II



TOP, KLM: From 2021c.
- But problems in TOP software happened and caused down time of operation.
  - Moved back to COPPER for a while.
- Hence, run coordinators make some rules for the PCIe40 upgrade to avoid potential problems affecting Belle II data taking.

ARICH: From 2022ab.
- Very stable without any major down time.

- **TRG, ECL, CDC, SVD:**
  - Replaced in 2022 summer and under validation.
  - Will be fully ready in LS1 (before 2023 autumn).

- PXD: Not in the present PCIe40 upgrade project due to its system design.
  - Discussion is ongoing for PXD to utilize PCIe40 for data collection to rescue slow pion in far future.

- Validation:
  - Detector local calibration run.
  - Cosmic run with < 500 Hz trigger.

  Check data quality, slow control, and compare with COPPER data
  - 30 kHz dummy trigger high rate test: Run for O(week).

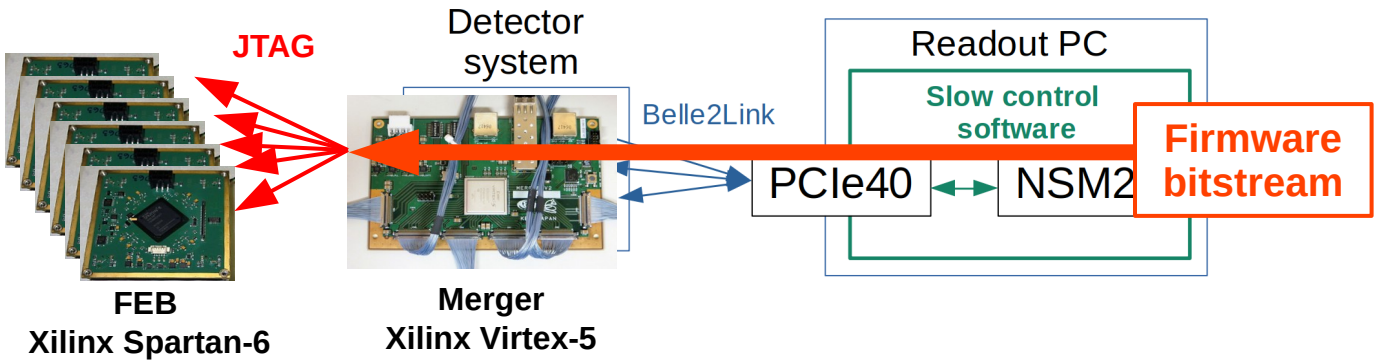
Upgrade specific for detector: ARICH

- ARICH system: 5~6 FEB → 1 Merger → Belle2Link → PCIe40
  - JTAG of FEB is controlled by Merger.

- Special Belle2Link slow control design: **Transferring an entire file**.
  - FEB firmware bitstream is transferred to Merger for each byte one-by-one.
  - Then Merger downloads the firmware to FEBs via JTAG and configure FEBs.

- Original Copper readout: 4 Mergers → 1 Copper. Each Merger processed one-by-one.
  - Consumed time: ~1.5 min.

- PCIe40 readout: 36 Mergers → 1 PCIe40. Parallel slow control processes.
  - The same consumed time: ~1.5 min.

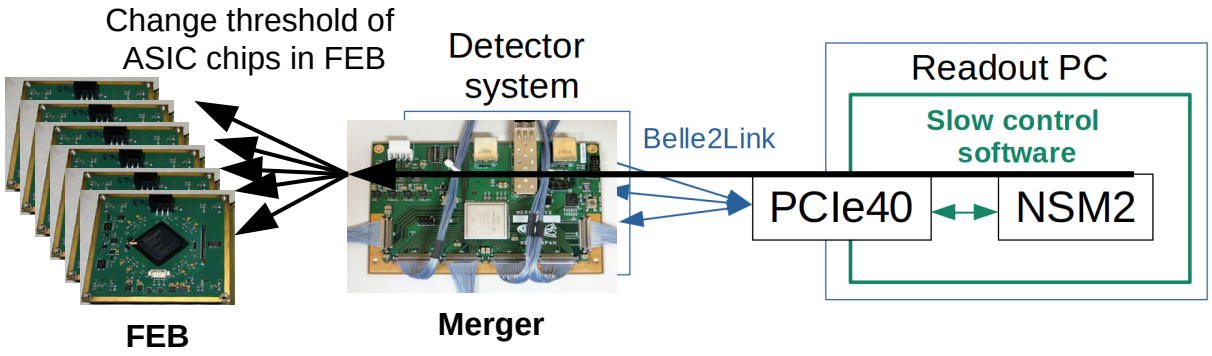


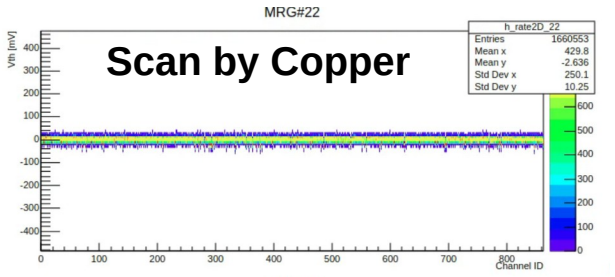
Validation for ARICH PCIe40 system: Local calibration run

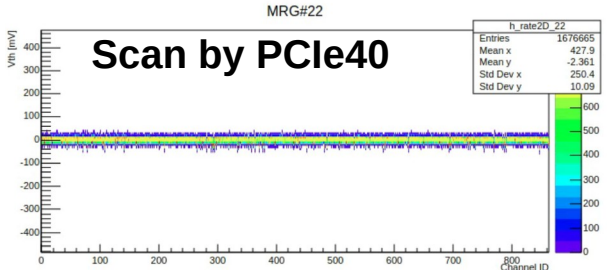
- For ARICH FEB, ASIC chips' threshold is controlled by slow control software in readout PC.

- Local calibration run for ARICH: threshold scan.
  - Hit rate per channel for each step of threshold value.

- Threshold scan as a validation for both:
  - Slow control: check how the ASICS configuration is correctly done by PCIe40 and software.
  - Data taking: check the data from Copper and PCIe40 are consistent or not.

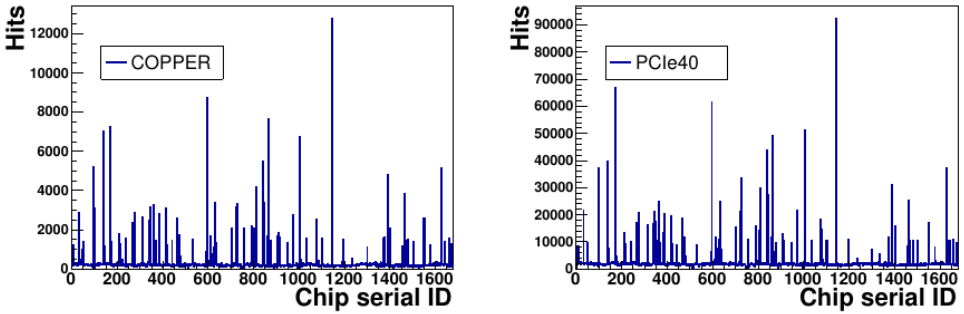




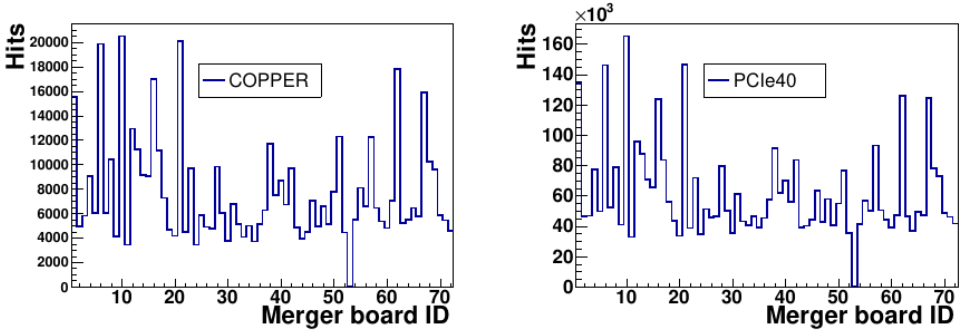


Validation for ARICH PCIe40 system: Cosmic run

- Cosmic run: Using cosmic trigger from L1 TRG system.
  - Data taking is done in global DAQ together with other detectors.
  - Meaningful physics data.
  - HLT: Online reconstruction, and Data Quality Monitor (DQM).
  - Check the DQM histograms to see if the COPPER and PCIe40 data are consistent.



(a) The number of hits in each ASIC chip



(b) The number of hits in each Merger

Upgrade specific for detector: TRG

- L1 TRG: Different from other detector systems using single type of FEE and its firmware.
  - 2 Universal Trigger boards: UT3 and UT4. Much larger FPGA than other FEE's.
  - 4 types of transceivers.
  - Several complicated firmwares for trigger logic.

- Difficulty: **Re-compilation of TRG UT firmware** after implementing new features.

TRG module	Board	Transceiver
2D tracker (x4)	UT4	UltraScale GTY
3D tracker (x4)	UT3	Virtex-6 GTH
Neural 3D tracker (x4)	UT3	Virtex-6 GTH
Event Timing Finder	UT4	UltraScale GTY
Track Segment Finder (x9)	UT4	UltraScale GTY, GTH
Global Reconstruction Logic	UT3	Virtex-6 GTX
Global Decision Logic	UT3	Virtex-6 GTX
TOP Trigger (x2)	UT3	Virtex-6 GTX

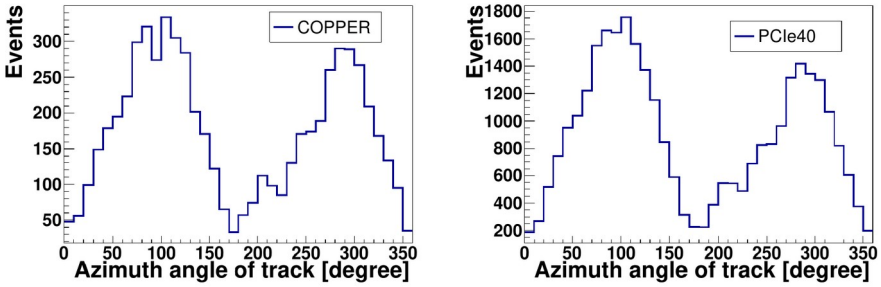
UT3
Xilinx Virtex-6
GTX, GTH

UT4
Xilinx UltraScale
GTH, GTY

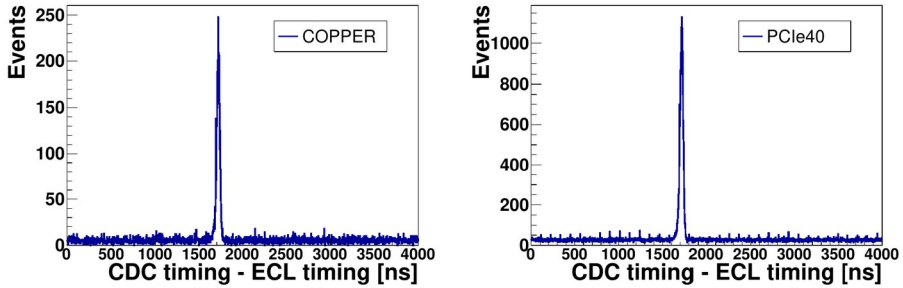
- Based on the difference of transceiver IPcore interface and the property:
  - Belle2Link and transceiver auto-reset scheme require adaption.
- Each TRG firmware needs to be updated, re-compiled, and validated one-by-one: Time-consuming.
- Update on all TRG modules' firmware have been complete in June 2022.

Validation for TRG PCIe40 system: Cosmic run

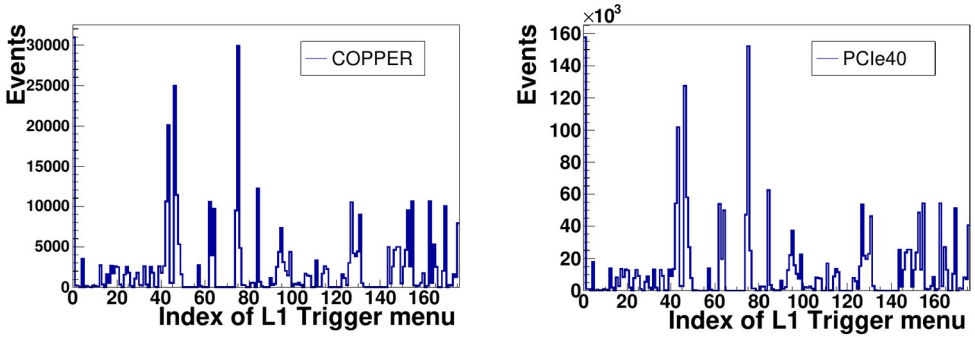
- For the validation of TRG system, we took cosmic runs with both copper and PCIe40.
  - Check if the DQM histograms from copper and PCIe40 are consistent.
  - Each TRG module has to be validated.



(a) The azimuth angle distribution from 2D trackers



(b) The event timing resolution by the difference between CDC timing and ECL timing

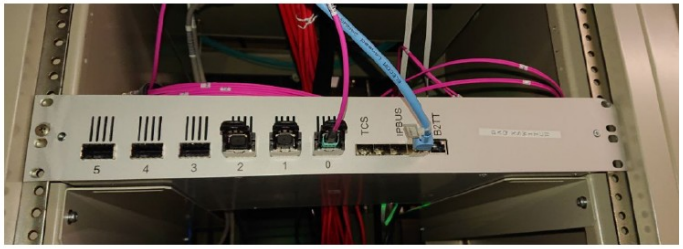


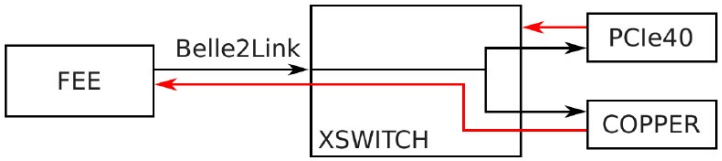
(c) The trigger rate of all the L1 final decision menu

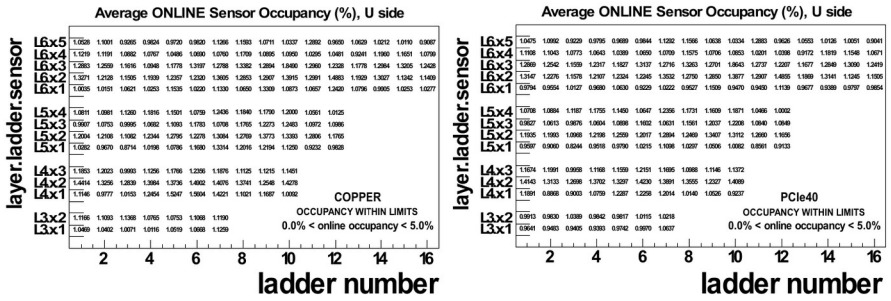
Upgrade specific for detector: SVD

- To test the PCIe40 system, we need to switch the optical cables every time, and resume the original connection after the test is done: Time-consuming.

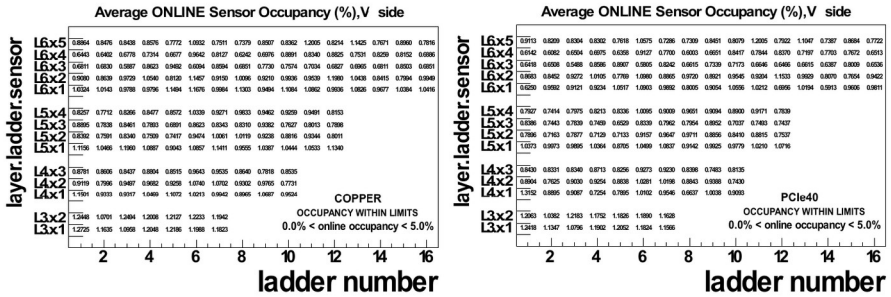
- For SVD, a **cross-point switch** is utilized:
  - Re-connection is not needed: Save lots of time.
  - PCIe40 test will not conflict with global run with COPPER.
    Even parallel data taking is fine.







(a) U side

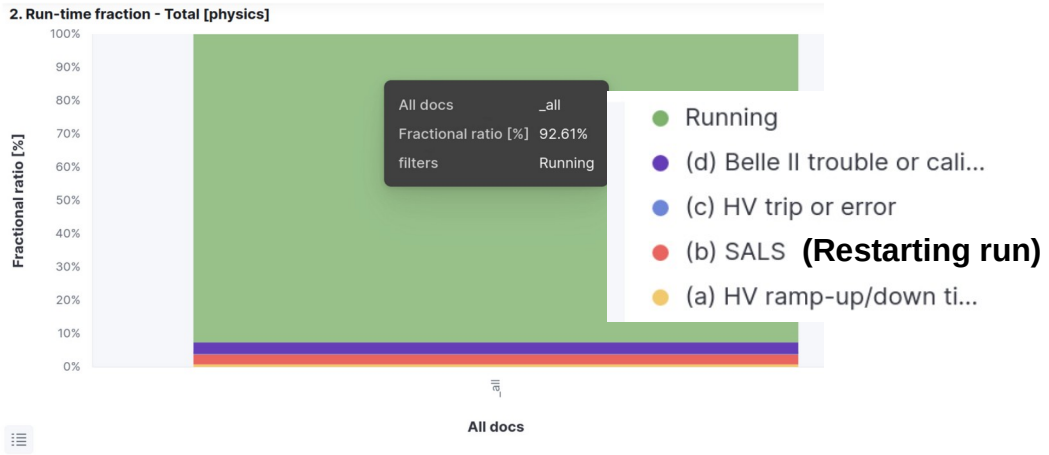


(b) V side

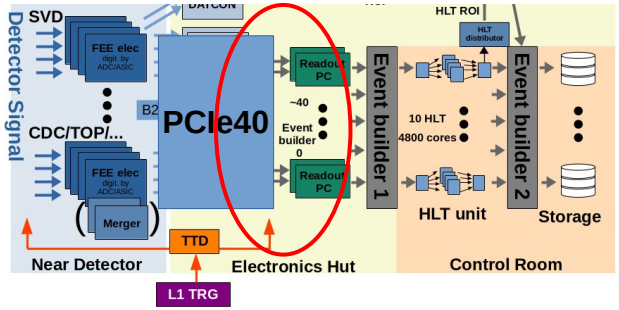
- Validation with cosmic run.
  - Consistent occupancy of sensors between COPPER and PCIe40 data.

Performance of the new system in 2022ab

- **Operation in 2022ab:**
  - Overall running time fraction in physics data taking: **92.6%**.
  - Restarting run: 3%.
  - System (detector or HV) problem: ~4%.
  - No major down time due to PCIe40.



- **PCIe40 → readout PC via PCI-Express:**
  - Using test pattern generator in PCIe40 firmware: ~3.9 GB/s.
  - Throughput in Belle II DAQ: 630 MB/s per readout PC.
  - Much improved from original Copper system using GbE (125 MB/s).



Performance of each system's PCI-Express readout in global run

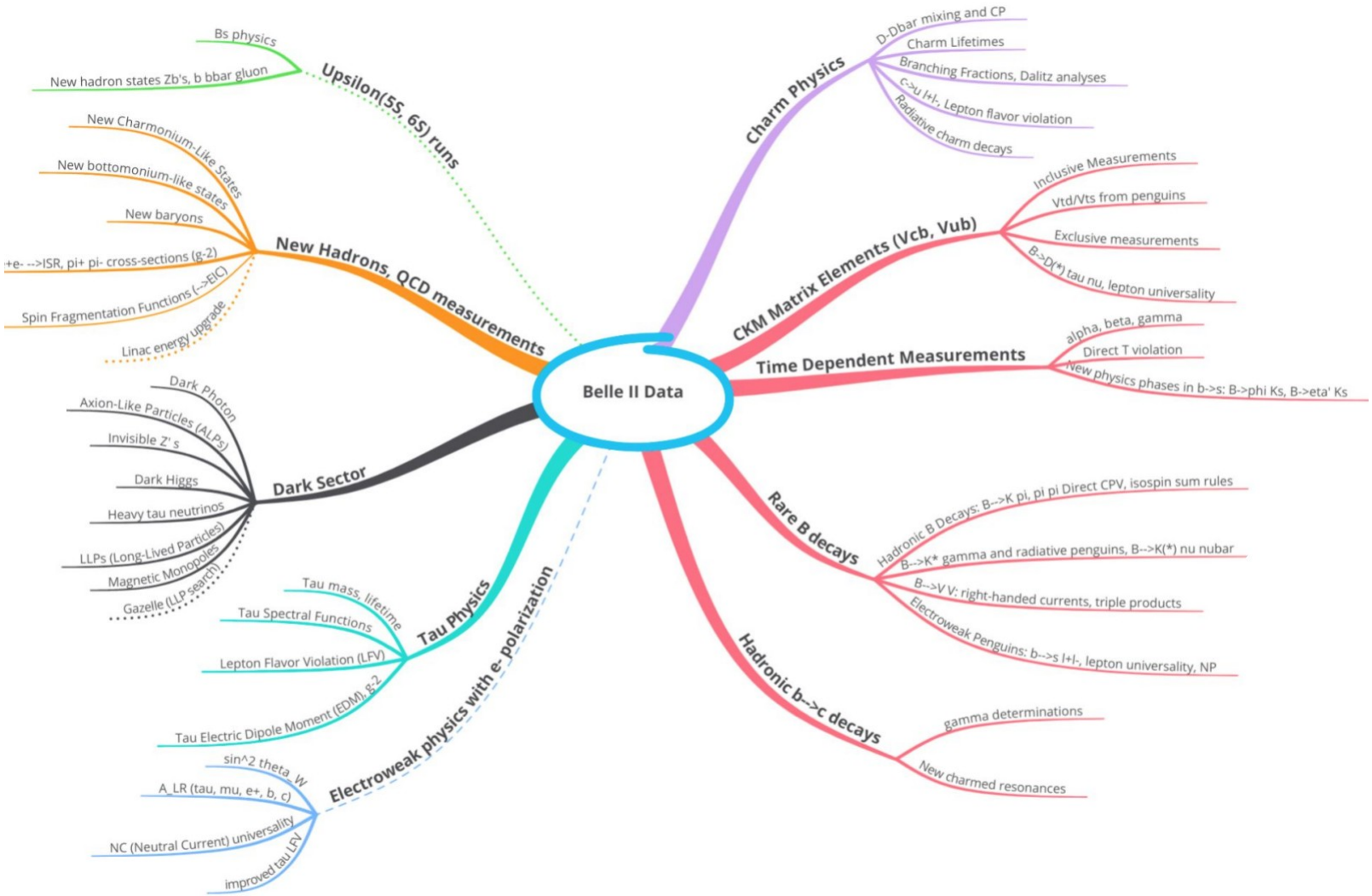
Detector	Max. Event size per readout PC (kB)	Max. throughput per readout PC (MB/s)	Run type
TOP	6.0	81.0	Physics
KLM	1.2	15.0	Physics
ARICH	5.9	86.2	Physics
TRG	0.94	26.0	Cosmic
	100.0	18.2	High rate test
CDC	1.5	0.26	Cosmic
	1.5	38.0	High rate test
ECL	4.3	0.8	Cosmic
	4.3	118.0	High rate test
SVD	7.3	1.3	Cosmic
	8.3	226.0	High rate test

Summary

- Belle II DAQ system adapts the PCIe40 board for the upgrade of readout.
- Development on PCIe40 firmware and software have been complete and validated including those specific for each sub-detector.
- In 2022ab physics data taking, TOP, KLM, and ARICH detectors have been running stable.
  - The rest of sub-detectors also finished the replacement in this summer, and their commissioning is ongoing.
- Commissioning with entire Belle II will be done in LS1 (up to autumn 2023).
  - Short-term plan for improvements: TTD link stability,
    double PCI-Express bandwidth, new event builder scheme, etc.
  - Long-term plan for PXD: utilize PCIe40 to rescue slow pion is also under discussion.

Backup

Belle II physics program

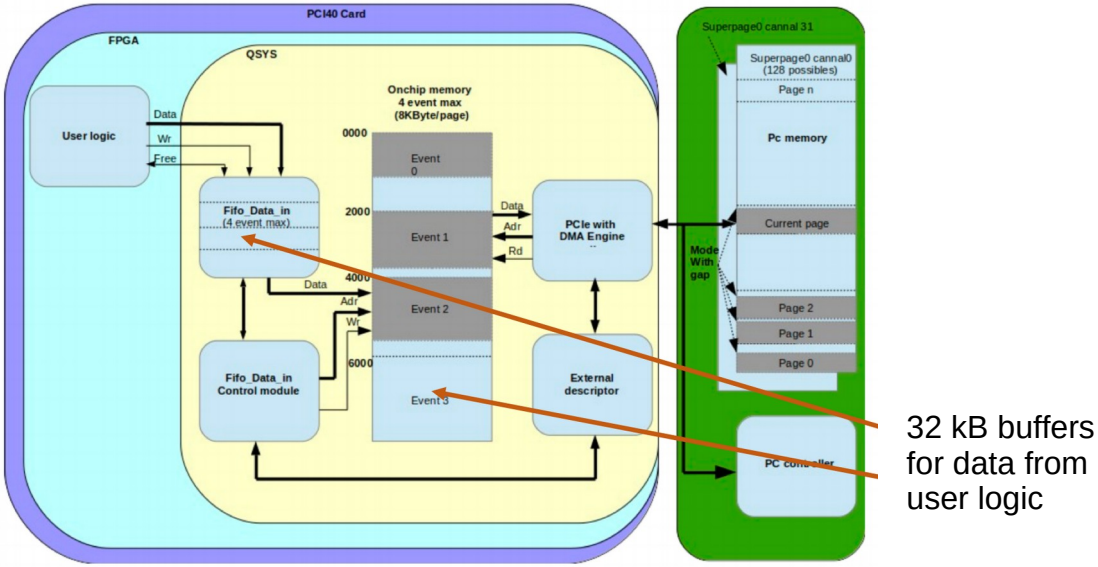


Number of readout devices for each detector

Detector	FEE with Belle2Link	COPPER	COPPER readout PC	PCIe40 with readout PC
SVD	52	48	9	5
CDC	300	75	9	7
TOP	64	16	3	2
ARICH	72	18	6	2
ECL	53	26	10	3
KLM	32	8	3	1
TRG	24	12	3	1

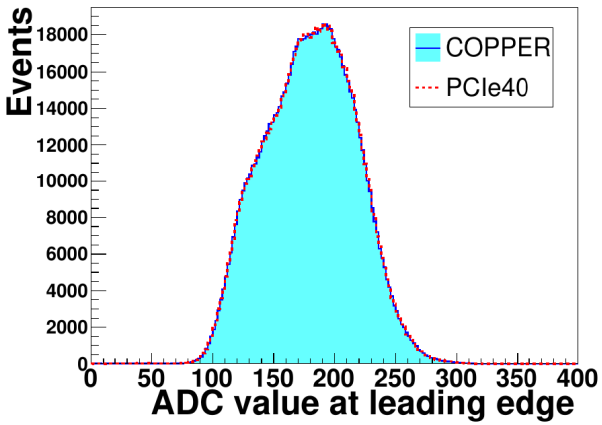
DMA interface

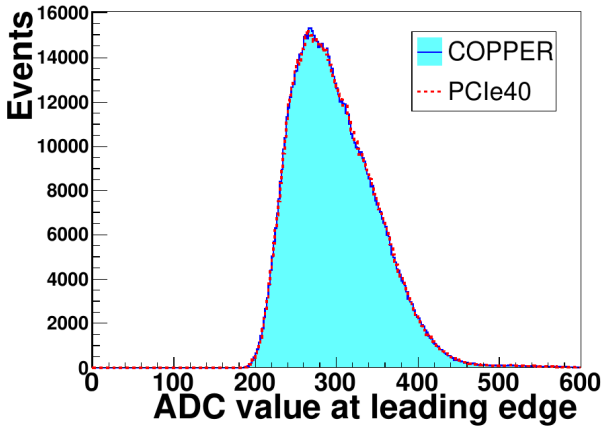
- Data transfer to readout PC: DAM interface with PCI-express.
- Implemented with ALICE DMA engine.
  - Intel FPGA IP core with external custom DMA controller.
- Well validated with test pattern generator and actual FE devices.
  - ~3.9 GB/s data transfer rate for PCIe x8 is reached.

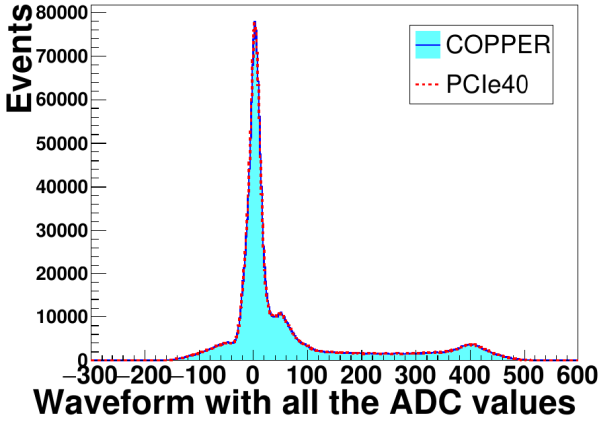


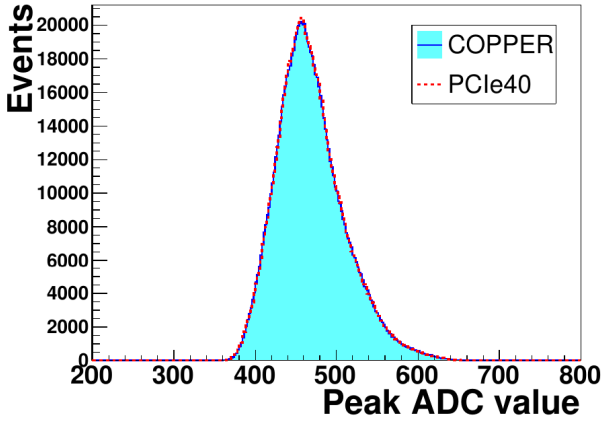
Validation for TOP PCIe40 system: Local calibration run

- Validation: local calibration run.
  - Test pulse signals are injected to the TOP FEE's ASIC channels with a frequency of 123.4 kHz
  - The data with 1M events are read out under a 1 kHz rate of dummy trigger.



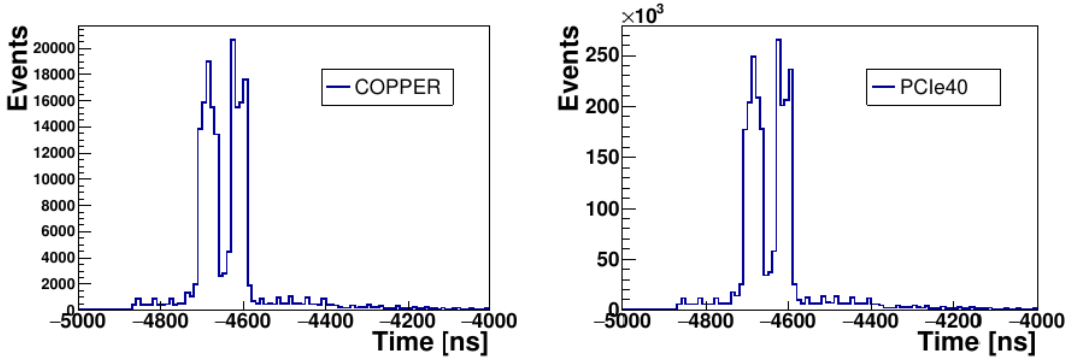




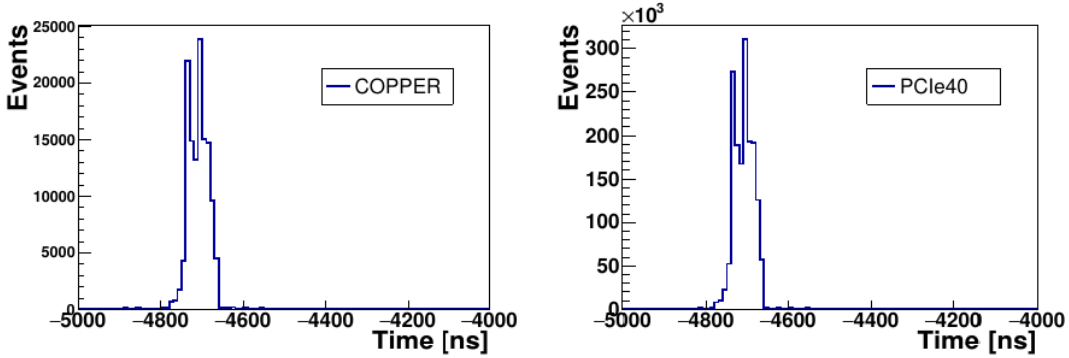


Validation for KLM PCIe40 system: Cosmic run

- Validation: Cosmic run.
  - Check the DQM histograms.



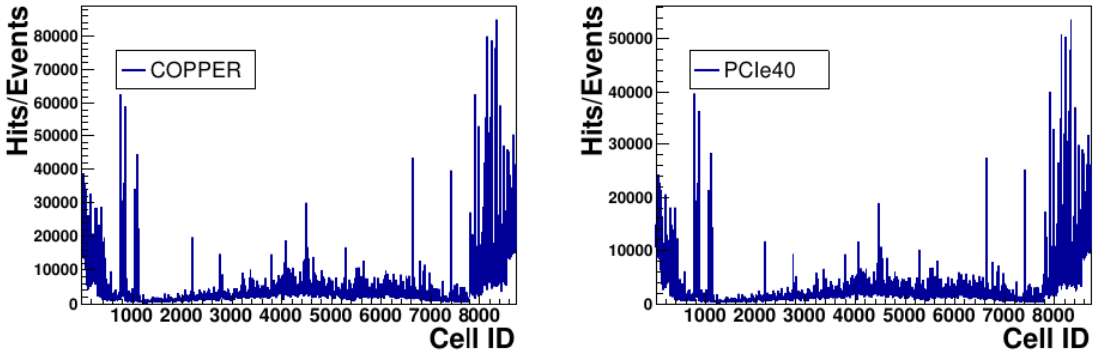
(a) The scintillator hit time in the barrel KLM



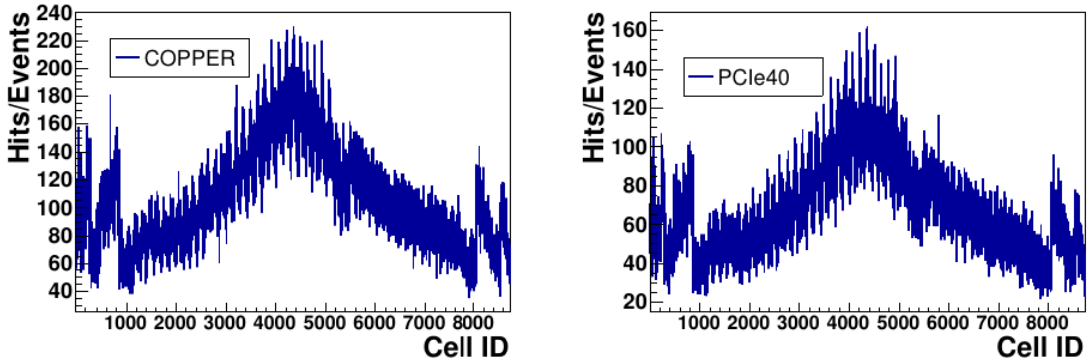
(b) The scintillator hit time in the end-cap KLM

Validation for ECL PCIe40 system: Cosmic run

- Validation: Cosmic run.
  - Check the DQM histograms.



(a) The fraction of hits of all the cells with no energy requirement



(b) The fraction of hits of all the cells where each cell is required to have an energy greater than 50 MeV