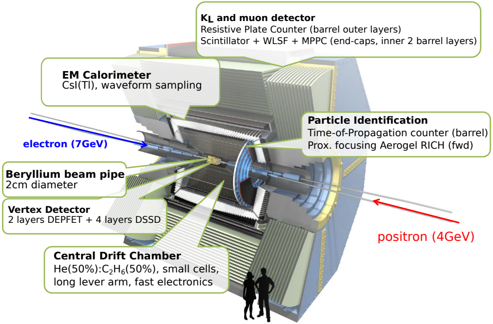
# Software-assisted Event Builder for the Belle II Experiment

Dima Levit

Institute of Particle and Nuclear Studies

April 9, 2024

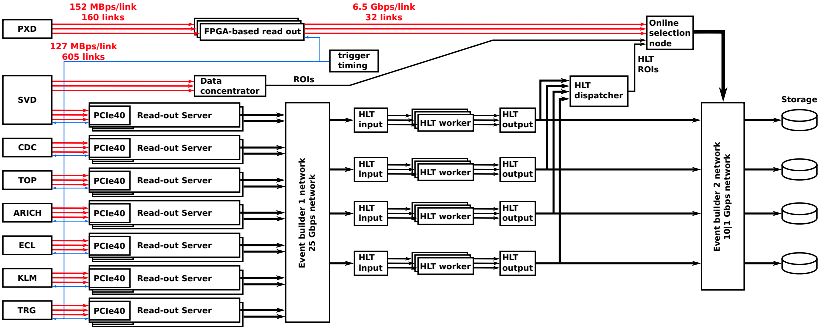
# Introduction



$K_L$ and muon detector
Resistive Plate Counter (barrel outer layers)
Scintillator + WLSF + MPPC (end-caps, inner 2 barrel layers)
EM Calorimeter
CsI(Tl), waveform sampling
Particle Identification
Time-of-Propagation counter (barrel)
Prox. focusing Aerogel RICH (fwd)
electron (7GeV)
Beryllium beam pipe
2cm diameter
Vertex Detector
2 layers DEPFET + 4 layers DSSD
Central Drift Chamber
He(50%):$C_2H_6$(50%), small cells,
long lever arm, fast electronics
positron (4GeV)

- Study of CP violation
- Search for physics beyond the SM
- Goal: 40x Belle data sample

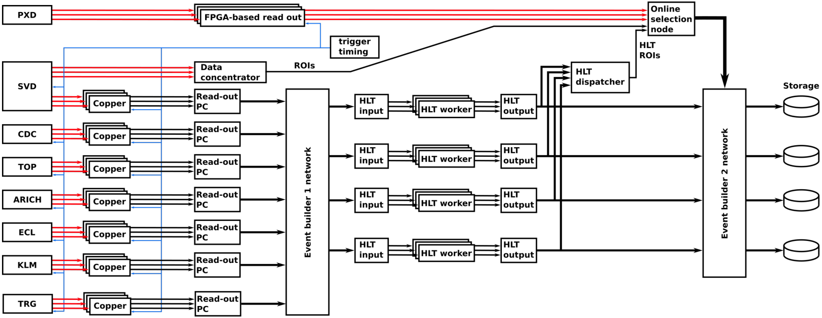
# Layout of the Belle II Data Acquisition System



- Separate read-out path for the pixel detector due to high data rate
- Two-stage event builder
- Online event reconstruction and data filtering
- Unified read-out system for 6 subsystems

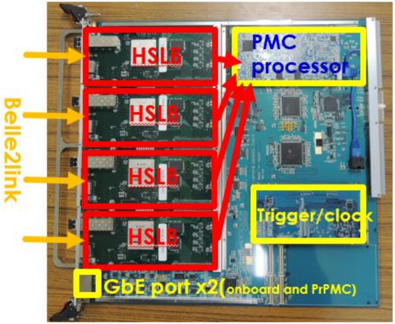
# Upgrade of the Read-out System of the Belle II Experiment

# COPPER Read-out System



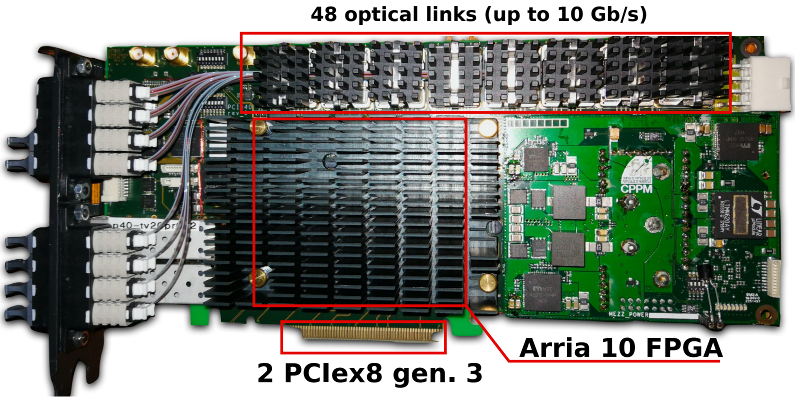
- COPPER: COmmon Pipelined Platform for Electronics Readout
- 600 read-out cards (HSLB), 150 COPPER modules

# COPPER System and Its Limitations



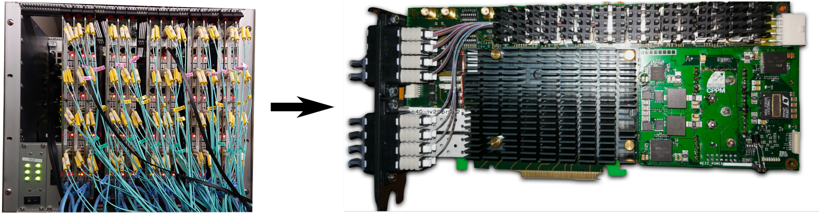
- Data receiver in Virtex-5 daughter boards
- Event building in Atom processor
  - 60 % CPU load at 30 kHz
- Output over 1 Gb/s Ethernet to read-out PC
- Expensive maintenance
  - deprecated hardware

# PCIe40 Hardware



- Designed for LHCb and ALICE
- Large Arria 10 FPGA
- No external memories

# COPPER and PCIe40 Comparison



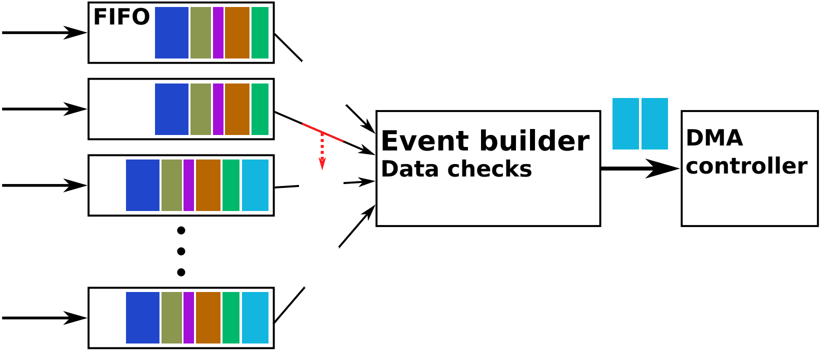
# Event Builder Algorithms

# System Requirements

- Designed trigger rate: 30 kHz
- Event size: up to few kB/channels
- 600 channels
- Fast event builder

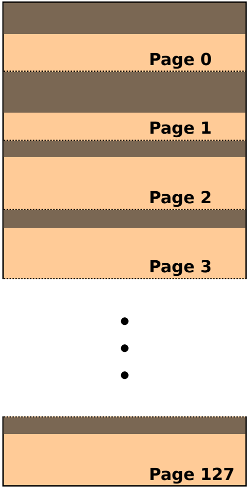
# Event Builder in Firmware

- Data buffering in internal memory
- Event builder
  - round-robin FIFO read-out
  - data consistency checks
  - self-triggered architecture
- Data read-out by DMA controller



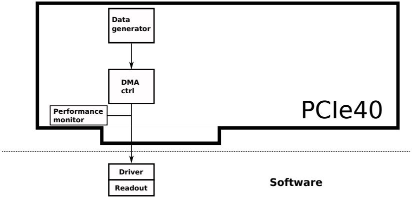
# DMA Controller and Memory Organization

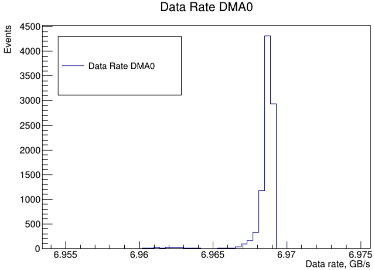
- Memory organization:
  - page-wise (8 kB) transactions: min. 1 page/event
  - superpage: 128 pages
- Ring buffer: 16 superpages in PC memory
- DMA operation:
  - superpage descriptors maintained by driver
  - page addresses calculated by DMA controller



# DMA Performance

- Setup
  - data generation in firmware at 250 MHz
  - discard data in software without processing
- Result: 6.965 GB/s

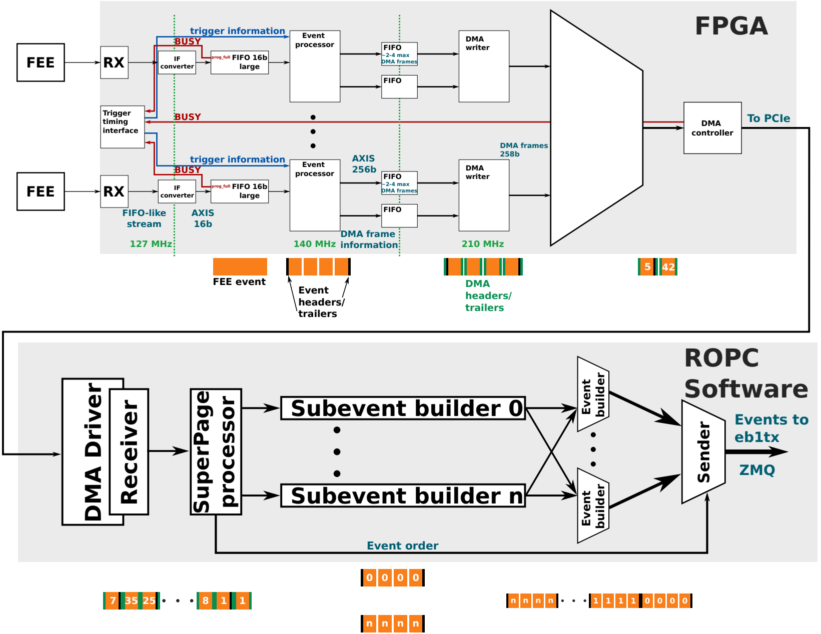




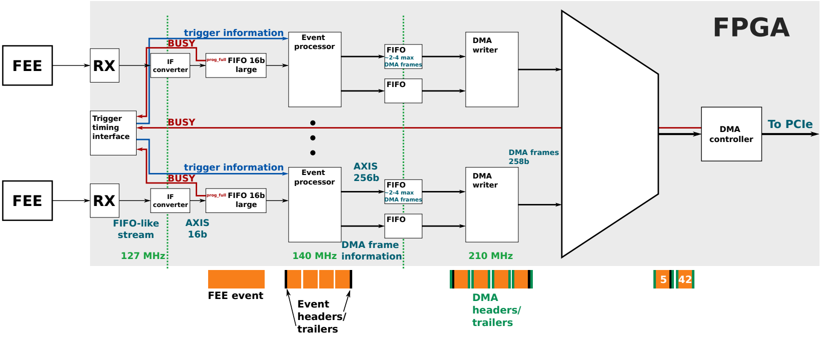
# Performance Limitations of the Firmware-based Event Builder

- System in operation in 2021-2022 run with 3 subdetectors
- High dead time
  - event buffering in on-board memory
  - round-robin event builder
  - large spread of event sizes
- Buffer overflow
  - late backpressure to trigger distribution
  - multiple events in FEE read-out chain
- Performance limited to 600 MB/s by software
  - inefficient CRC calculation
  - 1 GB/s without CRC calculation

# Software-Assisted Event Builder



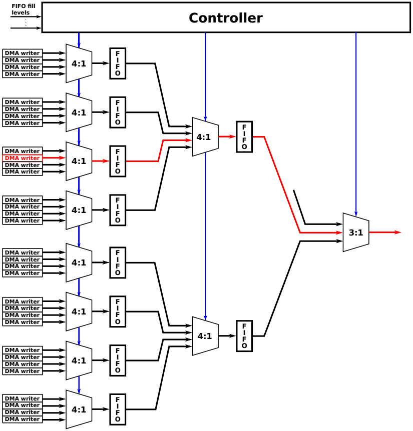
# Data Processing in Firmware



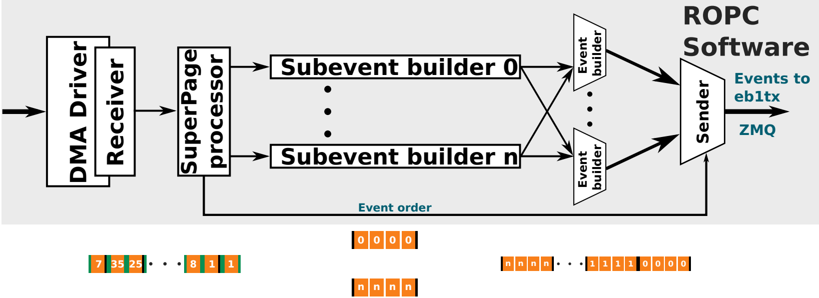
- Synchronization of trigger and data streams
- Data consistency check
- Event separation into fragments
- Routing of fragments to DMA controller
- Flow control based on DMA descriptor FIFO fill level

# Multiplexer

- Intelligent link readout scheduling
  - highest FIFO fill level
  - data availability
- Reevaluation on the frame-by-frame basis

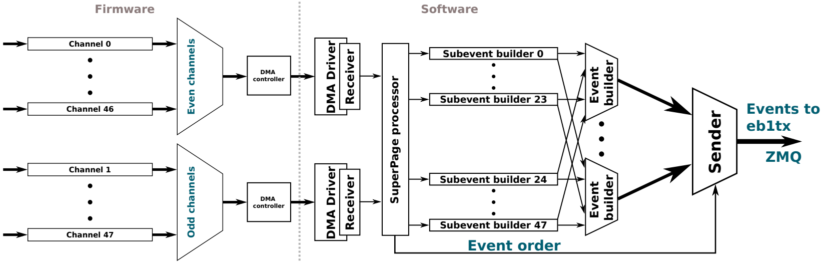


# Data Processing in Software

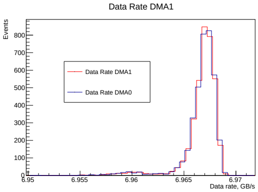


- Pulling events from driver
  - 1 CPU core reserved for receiver + driver
- Subevent reassembly
  - check of synchronization and data format
- Multiple event mergers to improve performance
  - CRC calculation
- Error recovery for CDC
  - high rate of SEUs in CDC
  - replace faulty event with a placeholder event
  - persistent until the next run

# Double PCIe Interface



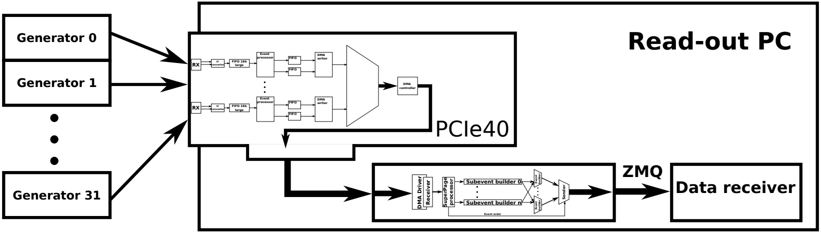
- Duplication of DMA controllers and receiver threads
- Almost no change to software or data processing logic
- Load balancing through separation of even and odd channels
- Throughput 14 GB/s



# System Performance

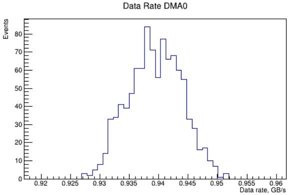
# Test Setup

- 32 data generators in HSLB
- Programmable event size distribution
- Data processing terminated in data receiver and data discarded
- ROPC configuration:
  - Intel(R) Xeon(R) Silver 4214R CPU (12 cores)
  - 24 GB memory, 3 channels (2400 MHz)



# Performance Measurements

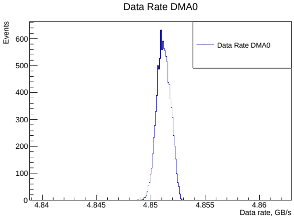
- Sarwate CRC calculation in event builder
  - 0.94 GB/s
  - 147.000 frames/s



# Performance Measurements

- Sarwate CRC calculation in event builder
  - 0.94 GB/s
  - 147.000 frames/s
- Slice-by-16* in event builder
  - 4.85 GB/s
  - 800.000 frames/s
  - limited by interface to DMA controller

*DOI: 10.1109/TC.2008.85

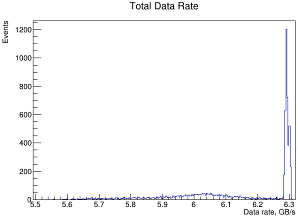


# Performance Bottlenecks

- CRC algorithm
  - solved by using Slice-by-16

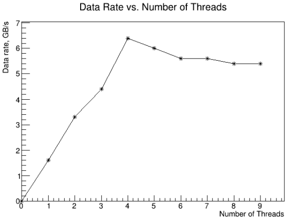
# Performance Bottlenecks

- CRC algorithm
    - solved by using Slice-by-16
- Interface to DMA controller
    - double number of PCIe interfaces
    - double PCIe performance: 6 GB/s
    - tail towards low data rates
        - ⇒ limitation of memory bandwidth



# Performance Bottlenecks

- CRC algorithm
  - solved by using Slice-by-16
- Interface to DMA controller
  - double number of PCIe interfaces
  - double PCIe performance: 6 GB/s
  - tail towards low data rates
    - ⇒ limitation of memory bandwidth
- Performance of the event builder thread
  - ∼1.56 GB/s/thread
  - scale number of threads to expected performance

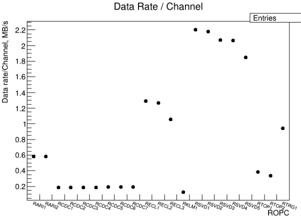


# Performance Bottlenecks

- CRC algorithm
  - solved by using Slice-by-16
- Interface to DMA controller
  - double number of PCIe interfaces
  - double PCIe performance: 6 GB/s
  - tail towards low data rates
    - ⇒ limitation of memory bandwidth
- Performance of the event builder thread
  - ~1.56 GB/s/thread
  - scale number of threads to expected performance
- Memory bandwidth
  - use faster memories
  - use more memory channels

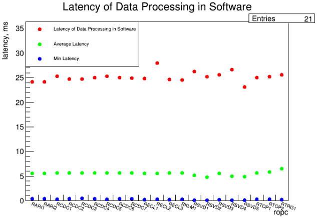
# Operation of the System in Belle II

- 21 systems, 600 detector links
- 1 DMA controller firmware
- Data taking with up to 3.6 kHz trigger rate
  - scales with luminosity
- Stable operation at 30 kHz artificial trigger rate without beams
  - 18.000.000 frames/s
- Data rates from 100 kB/s to 2.2 MB/s per channel
- Average event processing time in software consistent between detectors
  - dominated by OS effects
  - inversely proportional to event rate



# Operation of the System in Belle II

- 21 systems, 600 detector links
- 1 DMA controller firmware
- Data taking with up to 3.6 kHz trigger rate
  - scales with luminosity
- Stable operation at 30 kHz artificial trigger rate without beams
  - 18.000.000 frames/s
- Data rates from 100 kB/s to 2.2 MB/s per channel
- Average event processing time in software consistent between detectors
  - dominated by OS effects
  - inversely proportional to event rate



# Summary

- New read-out system for Belle II experiment
  - combined data processing in firmware and software
- System characterized and performance measured
- Throughput of 6 GB/s measured on the testbench
  - much higher than currently needed by Belle II
  - headroom for future detector upgrades
- Stable operation with Belle II detector at a fraction of the peak performance