Overview of Belle II DAQ system

R. Itoh, KEK

Livetime in 12hrs (rough estimation) = 94.0 % (94.0 % w/ all subsystems)
While HV permitted: 95.5 % (95.5 % w/ all subsystems)
1. Introduction to Belle II DAQ

Design Policy

- Conventional trigger-synchronized DAQ sequence.

- Deadtime-less design: pipelined trigger flow control.

- **COPPER-based readout**: combat proven scheme in Belle.

- Scalable back-end DAQ.

- **Unified Software Framework**
  * The same offline software framework (*basf2*) runs on every component in DAQ (even on COPPERs!)

- **“Non-Stop” DAQ**
  * Once DAQ is started, it is running all the time.
  * If a trouble occurs in detector FEE, stop the trigger distribution, fix the trouble locally, and restore the trigger without restarting other DAQ components.
Distribute L1 trigger and system clock to \(\sim 1000\) nodes

- **Fast control** — reset broadcast, partitioning, command to single subsystem, collect status info

- **Single type of module** — **FTSW** [Frontend Timing SWitch] with a few types of additional daughter cards
Data Flow in Belle II DAQ

- **FE dig**
- **PXD**
- **SVD**
- **CDC**
- **PID**
- **ECL**
- **Belle2link**
- **Rocket IO over fiber**
- **KLM**
- **Near detector**
- **E-hut**
- **DAQ server room**
- **DATCON**
- **~0.5M chan.**
- **~250 R/O boards (COPPERs)**
- **~40 R/O PCs**
- **R/O PC**
- **PXD readout box**
- **DATCON**
- **RAID ~10 units**
- **Express Recon**
- **basf2” on Linux CPUs**

- **~40 R/O PCs**
- **~O(10) units**
- **~400 cores/unit**
- **Event Builder 1**
- **Event Builder 2**
- **Reco**
- **DATCON**
- **Third**
- **Event Builder 2**
- **Reco**
- **RAID ~10 units**
- **Express Recon**
Data Flow in Belle II DAQ

- **PXD**
  - FE
dig
  - ~250 COPPERs
  - ~0.5M chan.
- **SVD**
  - FE
dig
  - tx
- **CDC**
  - FE
dig
  - tx
- **PID**
  - FE
dig
  - tx
- **ECL**
  - FE
dig
  - tx
- **Belle2link**
  - Rocket IO
    - over fiber
- **KLM**
  - FE
dig
  - tx

- **Near detector**
- **E-hut**
- **DAQ server room**
- **PXD readout box**
  - 30kHz
    - 1MB/ev
  - 10kHz
    - 100kB/ev
  - 1/10 size reduction
  - Feed back to PXD R/O
- **R/O PC**
  - 30kHz
    - 100kB/ev
  - 1/3 rate reduction
- **RAID**
  - ~10 units
- **Express Reco**
- **Event Builder 1**
  - HLT farms
    - ~O(10) units of
      - ~400 cores/unit
  - ~10kHz
    - 300kB/ev
    - = 3GB/s
  - 10kHz
    - 30kHz
    - 1MB/ev
    - 10kHz
    - 100kB/ev
    - 10kHz
    - 200kB/ev
  - 10kHz
    - 300kB/ev
COPPER and Belle2link

Belle2link Receiver (HSLB)

COPPER board

- CPU (Linux)
  - ATOM 1.6GHz
  - Memory: 512MB

- 1000Base-T port x 2

- Timing Receiver

- ~ 250 COPPERs
  - Network-booted SL5.5
  - basf2 runs on it.

- Dedicated readout software by Yamada-san
- In the FPGA on detector front-end card, "virtual" FINESSE is implemented, and it talks with "Belle2link transmitter core".
- In COPPER, Belle2link receiver (HSLB) is implemented instead of digitizer FINESSE, and connected to front-end card via optical fibers.
- The receiver "remote controls" the "virtual FINESSE" (slow control) and receives the data stream via optical fibers as if the remote FINESSE is implemented on the COPPER.

developed under collaboration with IHEP China (Zhen'An Liu's group)
Event Builder

Current design

- Each ROPC sends data via single GbE
- Layer2 switch and Layer3 are connected by **multiple** 10G
High Level Trigger (HLT)

- Unit structure (~10 units, 320 CPU cores/unit)
  * to reduce the number of output port of event builder
  * to keep up with the gradual increase of accelerator luminosity
  * fault-tolerant: each unit is completely independent

- Based on the parallel processing technology developed for basf2
Software Trigger Strategy on HLT

- **One unit** processes
  * 3-5kHz L1 rate (of total of 30kHz) with event size <100kB
  * 1/2 rate reduction with “Level 3” filter
    - Based on fast CDC tracking + ECL clustering.
    - Cut in the track |z| position and ECL energy sum
  * Full event reconstruction using all detector signals except PXD
  * Software trigger using physics event skim codes (Hadronic/tau event selection....) + Monitor trigger -> 1/3 reduction

Expected rate reduction : 1/3 = 1-2 kHz/unit at output.
- HLT performs special low momentum tracking and obtain "RoI" in PXD surface for reconstructed tracks.
- "RoI" is sent to PXD readout box for HLT-taken events.
- PXD box associate PXD-hits with RoI by FPGA processing and only associated hits are sent to 2\textsuperscript{nd} EVB.
- \( \rightarrow \) \( \sim \)1/10 reduction (data size) + 1/3 (rate) expected.
- Need to manage two different frameworks: NSM2 and EPICS.
- NSM2 is a home-grown slow control framework used in main DAQ components.
- EPICS is used in some of detector subsystems and SuperKEKB accelerator.
- A transparent environment is being developed.
Data Quality Monitoring

* based on real time histogram transport

HLT worker nodes

histogram collection from 3 servers

object transport over TCP

HLT control node

Express Reco

DQM server

DQM Browser
SuperKEKB/Belle II: Operation History

- Phase 1: Accelerator tuning / Vacuum scrubbing.

- Phase 2: Test run with outer detectors + pilot VXD detector

- Phase 3: Physics data taking with all (but not complete) detectors
Phase 2 DAQ

**Phase2 SVD**
- Dummy Trigger from master FTSW
- Timing Distribution
- not ready for high rates
- partially integrated and confirmed 30kHz

**Phase2 PXD**
- Pocket DHC
- limited up to ~10kHz

Full backend with 5 units of HLT+Storage

Online Storage

**Event Builder** 1

HLT 1
- HLT 2
- HLT 3
- HLT 4
- HLT 5

**Storage** 1
- Storage 2
- Storage 3
- Storage 4
- Storage 5

**Pocket ONSEN**

**EVB2**
Phase 3 DAQ

Timing Distribution

Beam Trigger from GDL

FADC/FTB

FEE

FEE

FEE

FEE

FEE

FEE

FEE

FEE

FEE

Full backend with 5 units of HLT+Storage (1600 cores of 6400)

QAS

Storage 1

Storage 2

Storage 3

Storage 4

Storage 5

Online Storage

Express Reco

ONSEN

EVB2

DHC

Phase3 PXD

HLT 1

HLT 2

HLT 3

HLT 4

HLT 5

Event Builder 1

FEE

FEE

FEE

FEE

FEE

FEE

FEE

FEE

FEE

Full SVD R/O

ECL

CDC

TOP

KLM

ARICH

TRG

COPPER

ROPC

QAS

Express

Reco

Phase 3 PXD
DAQ Status in Phase 3

- DAQ is basically running stably.

- Nominal L1 rate is around 3.5 kHz at $L = 5.5 \times 10^{33}$

- The overall DAQ efficiency is still 80 – 85 %. But when the beam and detector operation is stable, the efficiency is more than 90%.

- Injection veto distribution via FTSW is working stably.

**Sources of DAQ dead time:**

1) PXD BUSY
   * Bad modules not sending data
2) CDC ttlost/b2ldown
   * FE reprogramming required
3) TOP BUSY
   * Still firmware debugging.
4) ECL BUSY
   * Wave form readout
5) TRG BUSY/ttlost
   * Still firmware/software debugging.
Detailed analysis of errors (non-PXD)

- ERRORS detected by COPPER and ROPCs
  - SVD: mainly after receiving large events
  - CDC: belle2link seems unstable in some links
  - TOP: event # jumps. FW work by TOP experts is ongoing.
  - ARICH: Stable. B2link errors in several COPPERs at the same time.
  - ECL: b2link is stable. Sometimes no events arrives at some COPPERs
  - KLM: Stable.
  - TRG: Event # jump.

- Recovery of HSLB from large event errors
  - Investigation ongoing

- COPPER CPU freeze

- Event size
  - Currently, the event size is within the expectation.
    (except for SVD with larger occupancy.)
- ECL
  * Mostly due to the switching to the full wave form readout
    -> caused BUSY many times
  * If it runs in normal mode, ECL is quite stable.

[Comment on full wave form readout from DAQ]
* We assumed that the full wave form readout is limited for the calibration purpose only during injection veto trigger.
* But we very recently recognized that ECL group is planning to make it default in DAQ for the hadron ID.
* It is not included in the original DAQ design.
  -> The data size increase was found to be manageable and we agreed to switch to this option in normal DAQ
  Further debugging is in progress -> ECL talk

- Problems related to TOP
  * FTSW trouble -> lost 3 hours
  * Problem in database interface application (daqdbprovider)
    -> lost 3 hours
  => Still under the investigation.
Livetime in 12hrs (rough estimation) = 92.1% (92.1% w/ all subsystems)
While HV_permitted = 93.4% (93.4% w/ all subsystems)

Links: last 24hrs last 7days all period VXI test

Good day (June 2)
Bad day (May 25)

Livetime in 12hrs (rough estimation) = 76.6 % (75.9 % w/ all subsystems)
While HV permitted: 76.6 % (75.9 % w/ all subsystems)
Overall DAQ status in Phase 3 (by Yamada-san)

Last week livetime ratio while HV of detectors are permitted by accelerator.
- > **85% (80% with all sub-systems)**

Livetime in a week (rough estimation) = **78.7 % (65.7 % w/ all subsystems)**
While HV_permitted : **84.8 % (79.9 % w/ all subsystems)**

Dummy trigger runs with HV off during machine study etc.
- Further improvements in DHH/DHI firmware were made and the problems seem to be fixed.
ROI selection

ONSEN selects only pixels within given ROIs by HLT extrapolation and DATCON calculation

- Currently no ROI selection possible for all modules.
- ONSEN firmware design requires increasing module ID with single DHH required for matching IDs between data and ROIs.
- Possible solution:
  Change order in DHC firmware before sending to ONSEN (will be tested at beginning of summer shutdown)

- ROI selection for remaining modules verified with data of Local DAQ

Event selection not affected.

- RoI based data reduction was not available in Phase3.

* Some mistake in the ordering of module ID in DHH? 
  -> fixed by firmware update? or cabling?
Why reprogramming of FEE is required? Unrecoverable SEU?

CDC front end operation

- CDC is basically stable
  - run / fifoerr / timer / feer / semmbe / semcrc various errors
  - SALS doesn’t work and we need to re-program FPGA on FE
- Unpacker error: need to re-program FPGA on FE
- These issues can be fixed within 2-3 min just after excluding

- 7 FEs are masked during phase-3 physics run
  - #247,204,218 (b2llost): fixed once ~2 years ago and appear again
    - replacement of board didn’t work. it was fixed by swapping cat.7 cable
  - #37,193 (b2llost): new
  - #97, 115 (ttlost/crc error/b2link error): new. it was occurred when we had resumed operation with Belle solenoid turned ON

* Why reprogramming of FEE is required? Unrecoverable SEU?
• Highest priority on addressing problems that require masking large parts of detector (e.g., full 128-channel module) and/or have long recovery times.
  o PS Lockup:
    • Continuing to address errors seen in feature extraction (noncompliant data from carriers).
    • Adding SEU monitoring in PL to assess... desired feature anyway.
  o Event number mismatch / invalid event number:
    • Maybe some progress in simulation?
    • Really could use some ideas here!
    • Is our experience completely unique? Anything to be learned from other subdetectors?

- Still various bugs in firmware
  -> further effort during summer shutdown.
ECL: Waveform readout

Raw data are saved:

1-Part of events (1/1000) to monitor FPGA logic

2- Random trigger events for Overlay background

3- E>E_thresh (50 MeV) for the hadron/gamma separation

1 and 2 have no problems

3 was implemented later and indicates any problems for steady run. Increasing of data size for 5-10%.

in the case of background bursts it causes DAQ crash.

- The increase in the data size is not so much and within a manageable level.
  -> DAQ group agreed to go with waveform readout for ECL.
We did not have problems in phase 2 and we did not have problems before continuous injection. Burst produce huge number of energetic hits (several thousands).

If the burst event happens while the previous «raw data» event has not been readout buffer overwrites. It causes run crash. But often we have crash of ECL firmware. The last is ECL firmware bug and should be fixed.

V. Zhulnov will work with it in August.
Fix that firmware would not crash
Work to eliminate readout huge background events

Try to identify the huge background events by trigger?
Veto events with E_tot>20 GeV?
It will allow us to store data now!

Further debugging of firmware is required to be prepared for burst events.
Identification of background event by having veto trigger with E_tot>20 GeV
-> already talked to trigger group.
DAQ crash caused by TRG

- Busy
  - Recovered by SALS.
- Slow control stuck in NOTREADY or FATAL
  - Recover by restarting SC
- CDCTRG dataflow lost due to CDCFE
  - Recover by masking or reprogramming the FE
  - GDL stops generating L1

- At 10 kHz, TRG caused busy.
  - Didn't crash run, but trigger rate went down to ~ 2 kHz.
  - Didn't happen with GDL+GRL.

- Needs more debug to stabilize, especially firmware and SC.
- Sugiura-san is studying various correlation in FTSW and accelerator parameters.
- Very useful to understand the beam condition.

Trigger rate and luminosity

---

R. Sugiura
Relation between trigger rate and deadtime

The relation between trigger input rate and dead time is fitted by \( y = ax^b \), and extrapolated to 30kHz.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Extrapolation to 30kHz [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARICH</td>
<td>0.01</td>
</tr>
<tr>
<td>CDC</td>
<td>0.03</td>
</tr>
<tr>
<td>ECL</td>
<td>0.0006</td>
</tr>
<tr>
<td>KLM</td>
<td>0.006</td>
</tr>
<tr>
<td>SVD</td>
<td>0.0002</td>
</tr>
<tr>
<td>TOP</td>
<td>0.009</td>
</tr>
<tr>
<td>TRG</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
ttlost investigation

- monitor b2tt phase adjustment result
  - At every linkup, b2tt does phase scan to find a safe operation point of the clock/data phase difference
  - The result is not monitored so far, but it should be useful info

- special firmware to quantify the b2tt connection quality?
  - b2tt just gives 0 or 1 and hard to get the error rate
  - if the clock/data phase difference is manually fixed from remote and make a scan, an eye-diagram (or bath-tub) like plot can be generated
  - In this case, the data should be a random pattern (e.g. PRBS-31)

- another special firmware to test the connection quality?
  - since b2tt is based on 8b10b, the data may be well balanced
  - using PRBS-31 pattern may be a good test

- homework, possibly for this summer
Current event size on COPPER (including DAQ overhead)

<table>
<thead>
<tr>
<th>Event size/COPPER</th>
<th>max. rate w/o loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 kB</td>
<td>75kHz</td>
</tr>
<tr>
<td>2 kB</td>
<td>40kHz</td>
</tr>
<tr>
<td>3 kB</td>
<td>25kHz</td>
</tr>
<tr>
<td>4 kB</td>
<td>20kHz</td>
</tr>
</tbody>
</table>

S. Yamada
Current total event size is \( \sim 100\text{kB/ev} \)
Backend Processing (Event Building / HLT / Express Reco)

- Event building is stably working.

- At the very beginning of Phase 3, HLT was unstable because of the incomplete Linux signal handling at STOP/ABORT. => Fixed by Nils and now the operation is quite stable.

- The maximum processing rate with 5 HLT units is now >10kHz.

- Time to STOP/ABORT HLT has been reduced to 30 sec (from up to 5 min.) by new DQM histogram storage scheme.

- HLT selection has been tested and confirmed. Finally it was turned on and now stably working.

- ExpressReco is also stably working.

- The DQM scheme using HLT and ExpressReco is in a good shape. * Reference histograms are superimposed for the comparison. * Quality checking scheme has been established.
Plan for Phase3-Fall run

- Basically keep using the same DAQ configuration in fall run.

+ HLT reinforcement during summer shutdown:
  1) Use of cvmfs to share the same software among HLT units.
  2) Addition of 5 more HLT units.
     * 2 units will be operated in the current configuration as a margin.
     * 2 units will be used for the test of new ZeroMQ HLT framework.
       -> will be prepared to be compatible with existing framework
           and added in the global DAQ time to time for the test.
     * 1 unit will be used for the test of new framework on SL7.

with various improvements in all subsystems